Assessment of MERRA-2 Land Surface Hydrology Estimates

ROLF H. REICHLE,[a] CLARA S. DRAPER,[a,b] Q. LIU,[a,c] MANUELA GIROTTO,[a,b] SARITH P. P. MAHANAMA,[a,c] RANDAL D. KOSTER,[a] AND GABRIELLE J. M. DE LANNOY[d]

[a] *Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland*
[b] *GESTAR, Universities Space Research Association, Columbia, Maryland*
[c] *Science Systems and Applications, Lanham, Maryland*
[d] *Department of Earth and Environmental Sciences, KU Leuven, Heverlee, Belgium*

(Manuscript received 5 October 2016, in final form 12 December 2016)

ABSTRACT

The MERRA-2 atmospheric reanalysis product provides global, 1-hourly estimates of land surface conditions for 1980–present at ~50-km resolution. MERRA-2 uses observations-based precipitation to force the land (unlike its predecessor, MERRA). This paper evaluates MERRA-2 and MERRA land hydrology estimates, along with those of the land-only MERRA-Land and ERA-Interim/Land products, which also use observations-based precipitation. Overall, MERRA-2 land hydrology estimates are better than those of MERRA-Land and MERRA. A comparison against GRACE satellite observations of terrestrial water storage demonstrates clear improvements in MERRA-2 over MERRA in South America and Africa but also reflects known errors in the observations used to correct the MERRA-2 precipitation. Validation against in situ measurements from 220–320 stations in North America, Europe, and Australia shows that MERRA-2 and MERRA-Land have the highest surface and root zone soil moisture skill, slightly higher than that of ERA-Interim/Land and higher than that of MERRA (significantly for surface soil moisture). Snow amounts from MERRA-2 have lower bias and correlate better against reference data from the Canadian Meteorological Centre than do those of MERRA-Land and MERRA, with MERRA-2 skill roughly matching that of ERA-Interim/Land. Validation with MODIS satellite observations shows that MERRA-2 has a lower snow cover probability of detection and probability of false detection than MERRA, owing partly to MERRA-2's lower midwinter, midlatitude snow amounts and partly to MERRA-2's revised snow depletion curve parameter compared to MERRA. Finally, seasonal anomaly R values against naturalized streamflow measurements in the United States are, on balance, highest for MERRA-2 and ERA-Interim/Land, somewhat lower for MERRA-Land, and lower still for MERRA (significantly in four basins).

1. Introduction

Retrospective analysis (reanalysis) data products are based on the assimilation of a vast number of in situ and remote sensing observations into an atmospheric general circulation model (AGCM) and provide global, subdaily estimates of atmospheric and land surface conditions across several decades. The recent Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2; Gelaro et al. 2016, manuscript submitted to *J. Climate*), provides data beginning in 1980 and continuing to the present. MERRA-2 replaces and extends the original MERRA dataset (Rienecker et al. 2011), which ended in February 2016. The MERRA-2 horizontal resolution (~50 km) and the publication latency of a few weeks remain similar to those of MERRA. Recent upgrades in the MERRA-2 atmospheric assimilation system facilitate the use of newer satellite observations that could not be assimilated in the original MERRA system. Moreover, MERRA-2 benefits from advances in the Goddard Earth Observing System, version 5 (GEOS-5), AGCM.

Another new element in MERRA-2 is its use of observations-based precipitation data products to drive the land surface water budget (and aerosol wet deposition). Precipitation is the dominant driver of land surface hydrologic conditions. In most reanalysis systems, including the original MERRA, the precipitation seen by the land surface is generated by the system's AGCM following the assimilation of atmospheric observations. The model-generated precipitation is subject to significant errors in amounts and timing, which adversely impact the land surface hydrology estimates. To avoid this

Corresponding author e-mail: Rolf H. Reichle, rolf.reichle@nasa.gov

DOI: 10.1175/JCLI-D-16-0720.1

problem, the MERRA-2 model-generated precipitation is corrected with gauge- and satellite-based precipitation observations before reaching the surface. Observation-corrected precipitation was also used in MERRA-Land, an offline, land-only reanalysis product that supplements the original MERRA product (Reichle 2012). Because of the precipitation corrections and important land model updates, primarily to the rainfall interception parameterization, MERRA-Land provided significantly better land hydrology estimates than the original MERRA product (Reichle et al. 2011).

The precipitation corrections in MERRA-2 are a refined version of those used in MERRA-Land [section 2a(1)]. Reichle et al. (2017) show that overall the (corrected) MERRA-2 precipitation has lower bias and higher correlation and anomaly correlation values versus a reference dataset compared to both the MERRA-Land precipitation and the (uncorrected) AGCM-generated precipitation from MERRA-2 and MERRA. Moreover, in MERRA-2 the precipitation is corrected within the coupled atmosphere–land modeling system, allowing the near-surface air temperature and humidity to respond to the improved precipitation forcing. MERRA-2 thus provides more self-consistent surface meteorological data than were used for MERRA-Land (Reichle et al. 2017). This enhanced self-consistency in the forcing data also contributes to improvements in the MERRA-2 land surface estimates. Additional improvements in MERRA-2 include the seamless spinup of the land surface initial conditions in the low and midlatitudes, where precipitation observations of sufficient quality are available for use in the reanalysis (Reichle et al. 2017). Despite the improvements, however, the MERRA-2 corrected precipitation is still subject to errors in the observed precipitation inputs and the AGCM background precipitation (used to temporally and spatially disaggregate the observed precipitation), errors that ultimately impact the quality of the land surface hydrology estimates.

The land surface model and parameters of MERRA-2 very closely resemble those of MERRA-Land [section 2a(2)], thus carrying the model improvements from MERRA-Land into the coupled atmosphere–land MERRA-2 reanalysis. The objective of the present study is the assessment of the MERRA-2 land surface hydrology estimates (excluding glaciated surfaces). An in-depth assessment of the MERRA-2 land surface energy balance is left for future study. The MERRA-2 skill is compared to that of MERRA-Land, MERRA, and, where possible, ERA-Interim/Land, a land-only reanalysis dataset produced recently by the European Centre for Medium-Range Weather Forecasts (ECMWF). The paper is organized as follows. Section 2 provides a brief description of the MERRA-2 system and the data used in this study. Next, the MERRA-2 estimates of terrestrial water storage (section 3a), soil moisture (section 3b), snow (section 3c), streamflow (section 3d), and interception loss fraction (section 3e) are evaluated against independent data. Finally, section 4 provides a summary of the findings and conclusions.

2. Data

a. *The MERRA-2 data product and system*

1) Overview

The MERRA-2 reanalysis is produced by the National Aeronautics and Space Administration (NASA) Global Modeling and Assimilation Office (GMAO) using the GEOS-5.12.4 system (Bosilovich et al. 2015, 2016; Gelaro et al. 2016, manuscript submitted to *J. Climate*; http://gmao.gsfc.nasa.gov/reanalysis/MERRA-2). MERRA-2 replaces and extends the original MERRA reanalysis (Rienecker et al. 2011) and includes updates to the AGCM (Molod et al. 2012, 2015) and to the global statistical interpolation (GSI) atmospheric analysis scheme of Wu et al. (2002). In addition to the atmospheric in situ and remote sensing observations assimilated in MERRA, the MERRA-2 system also ingests observations from newer microwave sounders and hyperspectral infrared radiance instruments, as well as other new data types. Moreover, the system preserves the global water balance during the analysis (Takacs et al. 2016), which mitigates the water balance discontinuities seen in MERRA and other reanalysis products owing to observing system changes (Robertson et al. 2014). MERRA-2 further includes a comprehensive aerosol analysis (Randles et al. 2016) and a mass balance over glaciated land surfaces, including Greenland and Antarctica (Cullather et al. 2014). However, MERRA-2 does not include a land surface analysis (besides the precipitation corrections discussed next).

The MERRA-2 precipitation corrections algorithm is a refined version of that used in MERRA-Land as discussed in detail by Reichle et al. (2017); in particular, see their section 2 and their Figs. 1 and 2. MERRA-Land used the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) unified gauge-based analysis of global daily precipitation (CPCU) product (Xie et al. 2007; Chen et al. 2008; ftp://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/GAUGE_GLB), which is available with approximately 2-day latency. These gauge-based, daily, 0.5° CPCU data

TABLE 1. Key characteristics of the reanalysis datasets used here.

	MERRA-2	MERRA-Land	MERRA	ERA-Interim/Land
Domain	Global atmosphere–land	Global land	Global atmosphere–land	Global land
Year first published	2015	2012	2010	2014
Output period	January 1980–present[a]	January 1980–February 2016	January 1979–February 2016	January 1979–December 2010[b]
Output grid (latitude-by-longitude)	0.5° by 0.625°	0.5° by 0.667°	0.5° by 0.667°	0.75° by 0.75°
Output time step	Hourly	Hourly	Hourly	3 hourly (fluxes), 6 hourly (states)
Surface meteorological forcing data	MERRA-2	MERRA	MERRA	ERA-I
Precipitation observations used	0.5°, daily, gauge based (CPCU) and 2.5°, pentad, satellite/gauge based (CMAP[c])	0.5°, daily, gauge based (CPCU)	None	2.5°, monthly, satellite/gauge based (GPCPv2.1)
Reference	Gelaro et al. (2016, manuscript submitted to *J. Climate*)	Reichle et al. (2011)	Rienecker et al. (2011)	Balsamo et al. (2015)

[a] MERRA-2 continues to be updated to the present with latency on the order of weeks.
[b] The present paper uses also unpublished ERA-Interim/Land data for the period January 2011 to December 2014 [section 2b(6)].
[c] CMAP data are rescaled to the GPCPv2.1 climatology prior to use in MERRA-2 (Reichle et al. 2017).

were combined with hourly MERRA precipitation as a background to construct hourly precipitation forcing with daily totals matching those of the CPCU observations, separately for each grid cell.

The MERRA-2 corrected precipitation also uses the CPCU product but differs from that of MERRA-Land in three important ways. First, the precipitation corrections in MERRA-2 were implemented within the coupled atmosphere–land reanalysis system. Therefore, the MERRA-2 land surface forcings at any given time are impacted by the corrected precipitation prior to that time (via land–atmosphere feedback during the simulation). Second, the precipitation corrections in MERRA-2 do not extend to the high latitudes, with linear tapering of the corrections between 42.5° and 62.5° latitude and no corrections poleward of 62.5° latitude (see Fig. 2 of Reichle et al. 2017). Third, over Africa the MERRA-2 precipitation corrections use the coarser 2.5°, pentad CPC Merged Analysis of Precipitation (CMAP) product (Xie and Arkin 1997; ftp://ftp.cpc.ncep.noaa.gov/precip/cmap/), which is based on satellite as well as gauge observations and available with approximately 7-day latency. The latter two changes were made because the poor quality of the gauge-only CPCU product in Africa and at high latitudes had a detrimental impact on MERRA-Land estimates, as will be shown below.

Table 1 summarizes key features of the MERRA data products. MERRA-2 covers the period from 1980 to the present and continues to be updated with latency on the order of weeks. The MERRA-2 AGCM is run on a cube–sphere grid with an approximate resolution of 50 km, the atmospheric analysis is conducted on a Gaussian grid of similar resolution, and the MERRA-2 output fields are interpolated to a 0.5°-by-0.625° regular latitude–longitude grid for publication. Note that the longitudinal resolution of the outputs is slightly different between MERRA-2 (0.625°) and MERRA/MERRA-Land (0.667°). Estimates of surface meteorological and land surface fields are available at hourly time steps. The specific MERRA-2 data used here include monthly total land water storage, snow mass, snow cover fraction (SCF), runoff, evaporation, and precipitation (GMAO 2015a); hourly soil moisture, snow mass, snow depth, and snow cover fraction (GMAO 2015b); and time-invariant land fractions (GMAO 2015c).

2) LAND MODEL AND PARAMETERS

The land surface model used in all MERRA systems is the Catchment model (Koster et al. 2000), and the brief summary of the model provided in this paragraph parallels that of Reichle et al. (2011). The Catchment model explicitly addresses subgrid-scale soil moisture variability and its effect on runoff and evaporation. The basic computational element of the model is the hydrological catchment (or watershed). Within each element, the spatial variability of soil moisture is diagnosed at each time step from the bulk water prognostic variables and the statistics of the catchment topography. The vertical profile of soil moisture is given by the

TABLE 2. Key land surface parameter values, input datasets, and modeling schemes of the MERRA products.

Category	Dataset, scheme, or parameter	MERRA-2	MERRA-Land	MERRA	Additional information
Soils	Soil texture	NGDC (Reynolds et al. 2000)			De Lannoy et al. (2014); Mahanama et al. (2015)
	Soil hydraulic parameters, pedotransfer functions	Adapted from Cosby et al. (1984)			
	Depth to bedrock	GSWP-2 (Dirmeyer and Oki 2002)			
		Bilinearly interpolated	Original (1° grid)	Original (1° grid)	
	Minimum soil depth	1.34 m	1 m	1 m	
	Surface soil moisture layer thickness Δz_S	0.05 m	0.02 m	0.02 m	
	Root zone soil moisture layer thickness Δz_R	1 m	$0.75 \le \Delta z_R \le 1$ m	$0.75 \le \Delta z_R \le 1$ m	
	Vertical decay factor for saturated hydraulic conductivity (GNU)	1.0 m^{-1}	2.17 m^{-1}	2.17 m^{-1}	
Snow	Minimum SWE in snow-covered area fraction (WEMIN)	26 kg m^{-2}	26 kg m^{-2}	13 kg m^{-2}	Reichle et al. (2011)
	Maximum snow depth of uppermost snow layer (DZ1MAX)	0.08 m	0.08 m	0.05 m	
Rainfall interception	Capacity of canopy interception reservoir (SATCAP)	LAI × 0.2 kg m^{-2}	LAI × 0.2 kg m^{-2}	LAI × 1.0 kg m^{-2}	Reichle et al. (2011)
	Areal fraction of canopy leaves onto which large-scale precipitation falls (FWETL; dimensionless)	0.02	0.02	1.0	
	Areal fraction of canopy leaves onto which convective precipitation falls (FWETC; dimensionless)	0.02	0.02	0.2	
Boundary conditions	Albedo	MODIS climatology (Moody et al. 2008)			Mahanama et al. 2015
		8 day (2000–11)	Monthly (2000–04)	Monthly (2000–04)	
	LAI, greenness fraction	Monthly climatology based on AVHRR NDVI (1982–98, 1°; Dirmeyer and Oki 2002)			
	Land cover	Global land cover characteristics database, version 2.0 (USGS 2000)			
	Topography	HYDRO1k (USGS 1996)			
Atmospheric surface layer	Surface turbulence scheme	Monin–Obhukov similarity (Helfand and Schubert 1995)	Louis (1979)	Louis (1979)	Molod et al. (2012, 2015)

equilibrium soil moisture profile and the deviations from the equilibrium profile, described by variables in a thin surface layer and a thicker ''root zone'' layer. Furthermore, the water table depth of shallow (unconfined) groundwater can also be diagnosed from the model prognostic variables (Koster et al. 2000, their Fig. 2). Moreover, in each element the evolution of snow water equivalent (SWE; or snow mass), snow depth, and snow heat content in response to surface meteorological conditions and snow compaction is modeled using three layers (Stieglitz et al. 2001).

Table 2 summarizes key land surface model input datasets, parameters, and modeling schemes, with a focus on the elements that are different between the MERRA data products. For MERRA-2, the Catchment model of the original MERRA system has been updated with the rainfall interception and snow model parameters of the MERRA-Land version (see also Table 2 of Reichle et al. 2011) and with revised soil parameters (labeled ''BLM2'' in De Lannoy et al. 2014). The rainfall interception parameter changes act to reduce the amount of rainfall that directly reevaporates from the canopy (section 3e). The most important snow model change concerns the minimum snow water equivalent (WEMIN) value, which governs the model's snow depletion curve, with the modeled SCF, given simply by SCF = SWE/WEMIN for SWE < WEMIN, and SCF = 1 otherwise, where SWE is the snow water equivalent in units of $\mathrm{kg\,m^{-2}}$. A larger WEMIN value thus implies lower SCF for a given amount of SWE under low snow conditions (section 3c).

The soil parameter changes (Table 2) include a revised surface soil moisture layer thickness $\Delta z_S = 0.05\,\mathrm{m}$ in MERRA-2 (compared to $\Delta z_S = 0.02\,\mathrm{m}$ in MERRA-Land and MERRA) and a fixed root zone soil moisture layer thickness Δz_R of 1 m (compared to $0.75 \le \Delta z_R \le 1.0\,\mathrm{m}$, depending on local conditions, in MERRA-Land and MERRA). Further changes in MERRA-2 soil parameters (De Lannoy et al. 2014) include the spatial interpolation of the depth-to-bedrock parameter, a decrease of the vertical decay factor for the saturated hydraulic conductivity from 2.17 to $1.0\,\mathrm{m^{-1}}$, and a few minor processing changes. Note that there was an error in a fitting procedure used for one of the underlying topography-related functions in the Catchment model, which can result in excessive surface soil moisture variability and a lack of variability in root zone and total profile soil moisture. This error potentially affects the simulation of soil moisture in about 2% of all land surface elements (De Lannoy et al. 2014) and is present in all MERRA data products to date, including MERRA-2. The error will be corrected in future datasets.

The time step for the land model integration is 7.5 min in MERRA-2 and 20 min in MERRA-Land and MERRA. Land surface albedo parameters are from the Moderate Resolution Imaging Spectroradiometer (MODIS), with MERRA-2 using an 8-day climatology compared to the monthly climatology used in MERRA and MERRA-Land (Rienecker et al. 2011; Mahanama et al. 2015). The monthly climatological vegetation parameters (leaf area index and greenness) are based on Advanced Very High Resolution Radiometer observations and are the same for all MERRA data products. Finally, MERRA-2 uses the surface turbulence scheme of Helfand and Schubert (1995), whereas MERRA and MERRA-Land used the scheme of Louis (1979).

b. *Evaluation data and approach*

The evaluation of the MERRA products rests on comparisons to independent datasets. This subsection summarizes the datasets, processing, and metrics used to evaluate the MERRA data products, including a brief description of the ERA-Interim/Land product for which skill metrics are computed for reference where possible. Metrics used here include the bias, the unbiased RMSE (ubRMSE; or standard deviation of the error), and the time series correlation coefficient R, each measuring different aspects of a product's skill (Entekhabi et al. 2010). Depending on the variable under evaluation, we focus on the most informative metrics. Throughout the paper, we use the term ''anomalies'' when we refer to time series for which the mean seasonal cycle has been subtracted. The correlation coefficient of the anomaly time series is referred to as the ''anomaly R'' value; note that the anomaly R metric typically implies a more stringent assessment in that this metric does not measure (sometimes uninformative) skill from large and reliable seasonal variations. For a given evaluation, common masks and minimum data requirements were applied to all datasets prior to computing the anomalies or metrics.

1) TERRESTRIAL WATER STORAGE

The Gravity Recovery and Climate Experiment (GRACE) satellite mission provides global retrievals of monthly terrestrial water storage (TWS) changes, where TWS includes groundwater, soil moisture, snow, and surface water (Swenson and Wahr 2006; Wahr et al. 2006; Landerer and Swenson 2012). The data used in this study are the (unscaled) level-3 GRACE monthly 1°-by-1° land gridded product (version RL05 spherical harmonics) from the NASA Jet Propulsion Laboratory for the 13-yr period from 2003 to 2015 (http://GRACE.jpl.nasa.gov; Swenson 2012). The TWS retrievals are reasonably accurate (~10–30-mm-error standard deviation) but have coarse resolution in time (monthly) and space

TABLE 3. In situ soil moisture measurement networks used for validation. The numbers of sites for the anomaly *R* metric are given in parentheses (fewer sites are available because the computation of the anomalies requires a longer record than is needed for the computation of the other metrics).

Network	Number of sites		Depth of sensors used for validation (cm)		Maximum time period used
	Surface	Root zone	Surface	Root zone	
SCAN	135 (131)	106 (96)	5	5, 10, 20, 50	2002–14
USCRN	111 (108)	78 (75)	5	5, 10, 20, 50	2009–14
SMOSMANIA	21 (21)	20 (20)	5	5, 30	2010–14
OZNet-Murrumbidgee	53 (29)	31 (29)	4	4, 45	2002–14
All networks	320 (289)	235 (220)	—	—	—

(~300–400 km at midlatitudes) (Rowlands et al. 2005; Swenson et al. 2006). The retrievals, which are based on measurements of Earth's gravity field, are independent of the reanalysis data and can thus be used to evaluate the reanalysis estimates.

For the evaluation against GRACE observations, the monthly mean reanalysis TWS estimates were first smoothed on their native grid using a 300-km half-width Gaussian filter to match the resolution of the GRACE observations and were then regridded onto the 1°-by-1° grid of the GRACE data product. Skill metrics were computed using data for the period from 2003 to 2015. Anomaly time series were computed by subtracting the mean seasonal cycle from each dataset, where the mean seasonal cycle was computed by averaging, for a given calendar month, the data from all 13 years. Note that because GRACE observations only provide changes in TWS (with respect to a reference value) but not a long-term mean climatology, bias cannot be evaluated.

2) SOIL MOISTURE

For the validation of surface and root zone soil moisture across a variety of climatological conditions and land surface characteristics we used in situ measurements from the U.S. Department of Agriculture (USDA) Soil Climate Analysis Network (SCAN; Schaefer et al. 2007), the U.S. Climate Reference Network (USCRN; Bell et al. 2013; Diamond et al. 2013), the Soil Moisture Ocean Salinity Meteorological Automatic Network Integrated Application (SMOSMANIA) sites in southwestern France (Albergel et al. 2008; Dorigo et al. 2011), and the Oznet network in Australia's Murrumbidgee catchment (Smith et al. 2012). In total, we used measurements from 320 sites for surface soil moisture and 235 sites for root zone soil moisture. Table 3 lists the number of sites, measurement depths, and time periods used here for each network. Note that the multiyear evaluation periods end with 2014 because of the limited availability of the ERA-Interim/Land data [section 2b(6)].

Unless noted otherwise, the processing and extensive quality control of the in situ measurements and the computation of the metrics match that of De Lannoy and Reichle (2016); their Fig. 5 illustrates the locations of the sensors. After quality control of the hourly data, the in situ measurements were aggregated into daily averages, metrics were computed from the daily data for each site separately, and the results were then averaged using a spatial clustering algorithm. A nearest-neighbor approach was used to map the reanalysis data to the site locations. Times and locations where the soil was frozen or snow covered in the in situ measurements or the reanalysis data were excluded from the metrics computation. Daily soil moisture anomalies were computed by subtracting a seasonally varying climatology (computed by applying a running 31-day window to the multiyear average daily values). Confidence intervals were computed for each site individually with consideration of temporal autocorrelations and then spatially averaged using the above-mentioned clustering algorithm.

3) SNOW

Snow cover fraction estimates from the reanalysis products were validated using observations from MODIS on the NASA *Terra* spacecraft for the period from 1 June 2000 to 31 May 2015. The level-3, daily, 0.05° climate modeling grid MODIS snow cover fraction observations (MOD10C1; Hall et al. 2006) were first screened for cloud contamination (less than 20% cloud coverage), lake ice, nighttime conditions, and general data quality (confidence index value greater than 60). Next, the observations were aggregated to the MERRA-2 grid, provided observations were available to cover at least 30% of a given MERRA-2 grid cell and day. MERRA-Land and MERRA snow cover fraction estimates were interpolated to the MERRA-2 grid. MERRA-2 grid cells were excluded from the evaluation if they (i) had a land fraction of less than 80%, (ii) had fewer than 50 observations during the evaluation period, or (iii) never had a positive snow cover fraction in any of the MODIS or reanalysis datasets

TABLE 4. Characteristics of the basins examined in this study and skill (anomaly R) of 3-month smoothed streamflow.

Basin	River name	Station name	basin area km^2	Latitude °N	Longitude °W	Period	Anomaly R value MERRA-2	MERRA-Land	MERRA	ERA-Interim/ Land
1	Missouri	Hermann (including basins 2, 4, 7, and 10e)	1 353 275	38.71	92.75	1980–97	0.86	0.85	0.75	0.90
2	Missouri	Ft. Randall Dam (including basins 4, 7, and 10e)	682 465	43.07	98.55	1980–2009	0.69	0.66	0.69	0.63
3	Ohio	Metropolis	525 770	37.15	88.74	1980–2010	0.90	0.92	0.71	0.92
4	Missouri	Garrison Reservoir (including basins 7 and 10e)	469 826	47.39	101.39	1980–2009	0.70	0.67	0.72	0.65
5	Upper Mississippi	Grafton	443 660	38.90	90.30	1980–2010	0.88	0.84	0.56	0.87
6	Colorado	Lees Ferry (including basins 10a and 10d)	289 562	36.87	111.58	1980–2004	0.50	0.43	0.60	0.66
7	Missouri	Fort Peck Dam (including basin 10e)	149 070	48.04	106.36	1980–2009	0.71	0.61	0.75	0.71
8	Arkansas	Ralston	141 064	36.50	98.73	1980–2008	0.83	0.85	0.51	0.76
9	Arkansas-Red	Arthur City	115 335	33.88	95.50	1980–2001	0.84	0.80	0.42	0.86
10a	Green	Greendale	50 116	40.91	109.42	1980–2004	0.55	0.49	0.59	0.71
10b	Potomac	Point of Rocks	25 000	39.27	77.54	1980–96	0.94	0.94	0.83	0.89
10c	Sacramento	Bend Bridge	23 051	40.29	122.19	1980–2003	0.94	0.94	0.94	0.96
10d	Gunnison	Near Grand Junction	20 533	38.98	108.45	1980–2004	0.54	0.47	0.65	0.66
10e	Musselshel	Moseby	20 321	46.99	107.89	1980–2008	0.62	0.60	0.73	0.55
10f	Rio Puerco	Bernardo	19 036	34.41	106.85	1980–2007	0.28	0.27	0.15	0.26
10g	Yakima	Near Parker	9 479	46.50	120.44	1980–2008	0.81	0.72	0.64	0.80
10h	Tuolumne	La Grange Dam	4 337	37.67	120.44	1980–2003	0.72	0.68	0.67	0.85
10i	San Joaquin	Mokelunme Hill	1 863	38.31	120.72	1980–2003	0.73	0.72	0.73	0.78
Area-weighted average over small basins (10a to 10i)			—	—	—	—	0.66	0.63	0.66	0.70

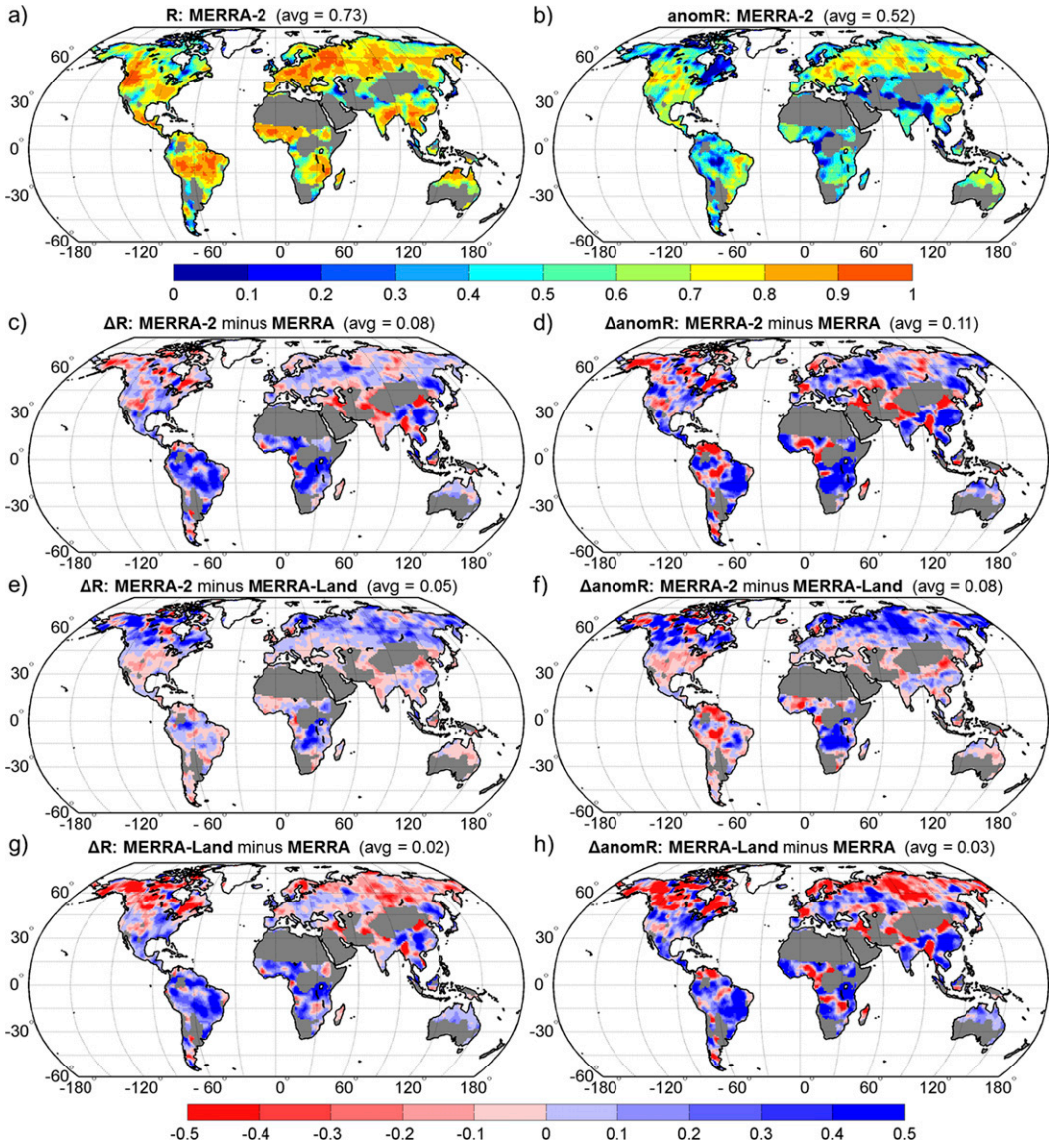

FIG. 1. (a) Skill of MERRA-2 TWS estimates, with skill measured as the monthly time series correlation coefficient R against GRACE observations for 2003 to 2015. (b) As in (a), but for anomaly R. (c) TWS skill difference ΔR for MERRA-2 skill minus MERRA skill. (d) As in (c), but for ΔanomR. (e) The ΔR for MERRA-2 skill minus MERRA-Land skill. (f) As in (e), but for ΔanomR. (g) The ΔR for MERRA-Land skill minus MERRA skill. (h) As in (g), but for ΔanomR. Titles indicate the spatial average across each map. Time series variability in masked (gray) areas is less than GRACE measurement error (section 3a).

at any time in the evaluation period. Note that MODIS snow cover observations rely on visible frequencies and are not available during polar night.

The reanalysis and observed snow cover fraction values on the MERRA-2 grid were then converted into binary snow cover estimates using a snow cover fraction threshold of 50%, and the probability of detection (POD) and probability of false detection (POFD; or false alarm rate) were computed from a contingency table (Wilks 2006, 260–265). The POD is the ratio of the number of correctly detected snow-covered conditions in the reanalysis to the number of snow-covered conditions in the observations, while the POFD represents the ratio of the number of erroneously detected snow-covered conditions in the reanalysis to the total number of snow-free conditions in the observations. Perfect scores are POD = 1 (higher is better) and POFD = 0 (lower is better).

Furthermore, the Canadian Meteorological Centre (CMC) snow analysis product (Brasnett 1999; Brown and Brasnett 2010) was used to validate snow amounts.

As summarized in Reichle et al. (2011), the CMC product provides daily snow depth throughout the Northern Hemisphere at a horizontal resolution of approximately 24 km for the period of March 1998 to the present. The product relies on optimal interpolation of in situ snow depth measurements and aviation reports with a first-guess field based on data from the Canadian forecast model. As in Reichle et al. (2011), we evaluated SWE estimates after converting the daily CMC product into SWE data using climatological snow density values. In addition, we also evaluated the skill in terms of snow depth (section 3c). Because of the limited availability of the CMC and ERA-Interim/Land data [section 2b(6)], we used only data for the period August 1998 to December 2014. For the evaluation, we regridded the reanalysis data onto the ~24-km, polar stereographic CMC grid. Daily SWE anomalies were computed by subtracting a seasonally varying climatology (computed by applying a running 31-day window to the multiyear average daily values). For each grid cell, days with zero snow in the CMC climatology are excluded from the computation of the skill metrics.

4) STREAMFLOW

We used naturalized, monthly streamflow gauge data (Livneh and Lettenmaier 2012; Mahanama et al. 2012) to assess the reanalysis runoff estimates. The naturalized streamflow data provide an estimate of the streamflow that would have occurred in the absence of anthropogenic hydrologic effects such as regulation at dams, evaporation from reservoir surfaces, and water withdrawals and return flows. The evaluation approach is the same as that of Reichle et al. (2011), except that here the comparison starts in 1980 instead of 1989 because we are no longer limited by the availability of ERA-Interim (hereafter ERA-I; Dee et al. 2011) data (which at that time was available only from 1989 onward). The data are for 18 basins in the United States, ranging in size from 1900 km^2 to 1 400 000 km^2, and the end dates for the comparison range from 1996 to 2010 depending on the basin (Table 4). Anomaly time series were computed by subtracting the mean seasonal cycle from each dataset, where the mean seasonal cycle was computed by averaging, for a given calendar month, the data from all years included in the comparison. Because the MERRA products lack routing schemes, we then applied a 3-month moving average to the (reanalysis and observed) anomalies prior to computing the time series anomaly correlation coefficients. Confidence intervals were computed with consideration of temporal autocorrelations (Reichle et al. 2011).

5) RAINFALL INTERCEPTION

The Global Land Evaporation Amsterdam Model, version 3.0a (GLEAM v3.0a), product provides daily global estimates of evaporation components, including rainfall interception loss, for the period 1980–2014 (Martens et al. 2016, manuscript submitted to *Geosci. Model. Dev. Discuss.*). Evaporation estimates are derived using a Priestley–Taylor model driven with satellite observations to the extent possible (Miralles et al. 2011). Rainfall forcing is from the Multi-Source Weighted-Ensemble Precipitation (MSWEP) product, which merges satellite, gauge, and reanalysis data (Beck et al. 2017). Interception dynamics are based on the model of Gash (1979), which was calibrated using a large number of in situ observations (Miralles et al. 2010).

Here, we investigate the interception loss fraction I, defined as the fraction of incoming rainfall (excluding solid precipitation) that is intercepted by the canopy and directly reevaporated back to the atmosphere without infiltrating the soil or contributing to surface runoff. For each product, we first computed the ratio of the annual total interception loss to that product's annual total rainfall. These annual I values were then averaged into climatological I values. Desert regions with annual mean MERRA-2 rainfall less than 0.2 mm day^{-1} were excluded from the analysis.

6) ERA-INTERIM/LAND

For soil moisture, snow amounts, and streamflow, we also compare the skill of the MERRA products to that of reanalysis estimates from ERA-Interim/Land (Table 1; Balsamo et al. 2015), a land-only reanalysis based on ERA-I. Like MERRA-Land, ERA-Interim/Land also uses a land surface model that was updated from the version of the underlying atmospheric reanalysis (ERA-I). Moreover, the surface meteorological forcing for ERA-Interim/Land is from ERA-I but corrected with monthly Global Precipitation Climatology Project (GPCP; Huffman et al. 2009) precipitation. [Note that in MERRA-2 and MERRA-Land, precipitation corrections are applied at daily or pentad time steps (Reichle et al. 2017).] Soil moisture in ERA-Interim/Land is modeled in four layers (0–7, 7–28, 28–100, and 100–289 cm). For the evaluation against in situ soil moisture measurements, we used the topmost layer as "surface" soil moisture and averaged the topmost three layers into a (0–100 cm) root zone soil moisture estimate (using weights proportional to each layer's thickness). As with the MERRA products, ERA-Interim/Land does not include a land surface analysis. However, ERA-I assimilates screen-level (2 m) observations of air temperature and humidity, which improves the surface meteorological estimates used to force ERA-Interim/Land.

Here, we use the 6-hourly ERA-Interim/Land data product that is publicly available at ~80-km resolution from 1989 to 2010 (http://www.ecmwf.int/en/research/climate-reanalysis/era-interim/land); this product used

GPCPv2.1 monthly precipitation data for corrections, except during 2010, when a climatological correction was applied. For 2011 to 2014, we obtained additional data from ECMWF that continue the public data record with monthly precipitation corrections based on GPCPv2.2 data (G. Balsamo and E. Dutra 2016, personal communication). An additional minor change in the 2011–14 ERA-Interim/Land product is the improved masking of ocean grid cells prior to interpolation from the original reduced Gaussian T255 grid onto the 0.75° output grid. We examined the impact of the ERA-Interim/Land version change on our results by repeating the soil moisture and snow amount evaluations (sections 3b and 3c) using reanalysis estimates through 2010 only (not shown). The results obtained using the shorter time period are very similar to those presented below and do not alter the conclusions. Note that the streamflow evaluation is only through 2010 [section 2b(4)].

3. Results

In this section, we evaluate several land surface water stores and fluxes from MERRA-2 and demonstrate consistent improvements compared to MERRA. For reference, estimates from MERRA-Land and, where possible, ERA-Interim/Land are included in the comparison.

a. Terrestrial water storage

GRACE TWS retrievals permit the validation of the total land water storage estimates (comprising groundwater, soil moisture, snow, and canopy water) from the reanalysis at the monthly, ~300–400-km scale [section 2b(1)]. Figure 1a shows the time series correlation coefficient of MERRA-2 TWS estimates versus GRACE retrievals. Averaged across the globe, the correlation coefficient is 0.73. In regions with a strong TWS seasonal cycle and/or significant interannual TWS variability (as indicated in the GRACE data), including the Pacific Northwest, the eastern United States, northern South America, the Sahel, and large portions of Eurasia, the correlation skill of MERRA-2 ranges from 0.7 to 1.0. Lower skill values are found in regions where the seasonal and interannual TWS variability is less pronounced, leading to signal-to-noise ratios that are commensurately smaller. Finally, in the masked (gray) areas the seasonal and interannual GRACE TWS variability is less than the latitude-dependent GRACE measurement error (Wahr et al. 2006, their Fig. 2), which is usually the case in deserts (Fig. 1a).

Next, Fig. 1c shows the difference in correlation skill ΔR between MERRA-2 and MERRA. Averaged across the globe, the MERRA-2 correlation skill exceeds that of MERRA by 0.08. The largest improvements are in South America and Africa, where the precipitation errors in MERRA vary greatly with season and where MERRA-2 perhaps benefits the most from the observation-based precipitation corrections (Reichle et al. 2017). In the northern high latitudes, the TWS correlation skill of MERRA-2 is, on average, similar to that of MERRA. Figure 1c further reveals a known issue with the CPCU precipitation product in Myanmar, where the TWS skill of MERRA-2 is considerably less than that of MERRA. Errors in the gauge measurements resulted in a large discontinuity in 2007 in Myanmar (Reichle et al. 2017). This precipitation discontinuity negatively impacts the local TWS time series in MERRA-2 (and MERRA-Land), leading to very poor correlation skill there.

Figure 1e shows the TWS correlation skill difference between MERRA-2 and MERRA-Land. Globally averaged, the MERRA-2 TWS skill exceeds that of MERRA-Land by 0.05. The biggest improvements in MERRA-2 over MERRA-Land are in Africa, where MERRA-2 benefits from corrections with the CMAP satellite-gauge precipitation product (as opposed to the gauge-only CPCU product used for MERRA-Land), and in the northern high latitudes, where MERRA-2 precipitation is not corrected while MERRA-Land precipitation is based on CPCU data. Again, since there are very few gauges in Africa and the northern high latitudes, the quality of the CPCU product in these regions was considered too poor for use in MERRA-2 (Reichle et al. 2017). The poor quality of the CPCU product in the northern high latitudes and in Myanmar is also reflected by the noticeably reduced TWS skill there for MERRA-Land compared to MERRA (Fig. 1g). Nevertheless, the TWS skill is higher for MERRA-Land than for MERRA in the Sahel and in southern Africa, despite the low rain gauge density there.

The skill of the TWS anomaly time series (for which the mean seasonal cycle is removed, thereby emphasizing interannual variations) is shown in Fig. 1b. The global average anomaly value for MERRA-2 is 0.52, with values of 0.7 and above in the northern midlatitudes and values around 0.5 in most other regions, except for the central Amazon basin, northwestern and eastern Canada (Yukon, Quebec, and Newfoundland), central Africa, and northern India and Myanmar, where the TWS anomaly R skill values are 0.3 and lower. In some cases the low skill is due to human-induced trends that are not modeled in the MERRA systems (e.g., the groundwater depletion in northern India; Rodell et al. 2009).

On average, the MERRA-2 anomaly R exceeds that of MERRA by 0.11, with strong improvements in many regions of the world (Fig. 1d), most notably in Africa. But there are also regions where the anomaly R skill of

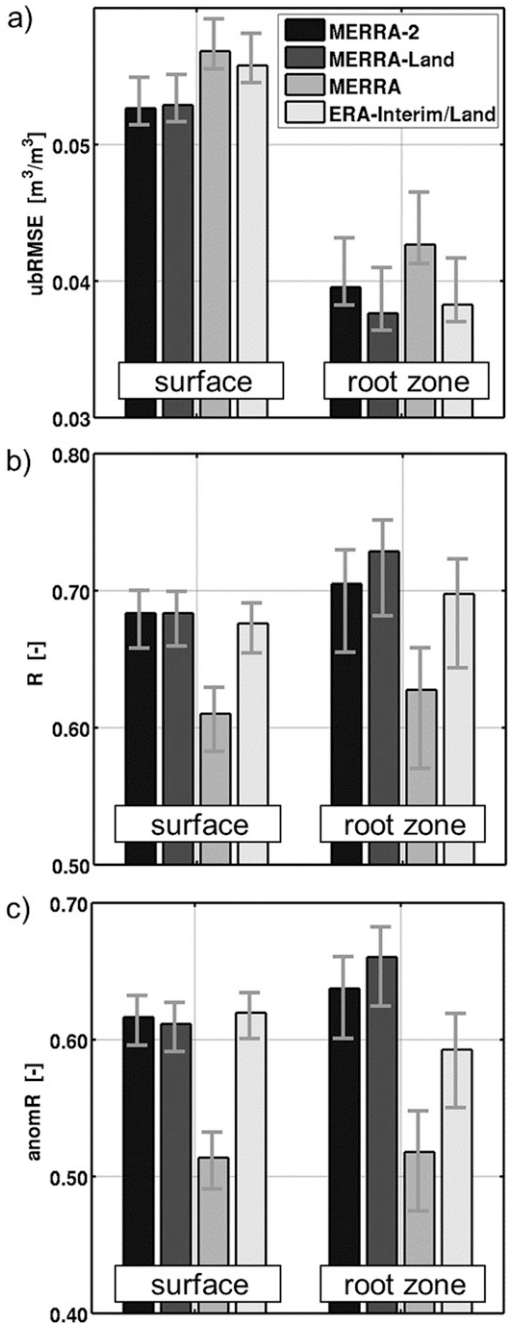

FIG. 2. Skill metrics against in situ soil moisture measurements: (a) unbiased RMS error (ubRMSE), (b) time series correlation coefficient R, and (c) anomaly time series correlation coefficient (anomR). Error bars show 95% confidence intervals. Metrics are cluster-based average skill values for MERRA-2, MERRA-Land, MERRA, and ERA-Interim/Land across stations from the SCAN, USCRN, Oznet, and SMOSMANIA networks (Table 3).

MERRA-2 is worse than that of MERRA, including portions of high-latitude North America, northern South America, West Africa, central Asia, and, naturally, Myanmar. MERRA-2 is also improved compared to MERRA-Land (average anomaly R difference of 0.08), with the strongest skill gains in the northern latitudes and Africa (Fig. 1f); this further supports the decision not to correct the precipitation in MERRA-2 at high latitudes and to use the coarser-scale but higher-quality CMAP precipitation product over Africa rather than the gauge-only CPCU data. Finally, the MERRA-Land TWS (anomaly R) skill is better than that of MERRA by 0.03 on average, with improvements in South America but widespread degradation in the northern high latitudes, again owing to the scarcity of precipitation gauges in this region (Fig. 1h).

The results for the TWS skill in terms of the ubRMSE (not shown) are generally similar to those for R and anomaly R, with improvements in MERRA-2 over MERRA and MERRA-Land primarily in South America and Africa. However, the MERRA-2 ubRMSE skill is similar to that of MERRA-Land in high-latitude Eurasia (not shown), despite the fact that the MERRA-2 R and anomaly R values are markedly improved there (Figs. 1e,f). This suggests that the amplitude of TWS variations in this region is worse in MERRA-2 than in MERRA-Land, which is most likely due to the excessive spring and summer precipitation in MERRA-2 in that region (Reichle et al. 2017).

b. *Soil moisture*

Figure 2 shows the skill of surface and root zone soil moisture estimates from MERRA-2, MERRA-Land, MERRA, and ERA-Interim/Land, measured against in situ measurements from the SCAN, USCRN, OZNet, and SMOSMANIA networks [section 2b(2)]. The average ubRMSE values range between 0.05 and 0.06 $m^3\,m^{-3}$ for surface soil moisture and around 0.04 $m^3\,m^{-3}$ for root zone soil moisture (Fig. 2a). The surface soil moisture ubRMSE for MERRA-2 and MERRA-Land is 0.053 $m^3\,m^{-3}$, which is significantly lower than the MERRA ubRMSE of 0.057 $m^3\,m^{-3}$, reflecting the positive impact of the precipitation corrections and the updates to the canopy interception reservoir and other model parameters in MERRA-2 and MERRA-Land [section 2a(2)]. The relative performance of the MERRA products is similar for root zone soil moisture, although here the difference between the MERRA-2 and MERRA ubRMSE is not statistically significant as indicated by the overlapping 95% confidence intervals.

The ubRMSE metric includes differences in the amplitude of the time series variations, which are partly a reflection of the scale discrepancy between the point scale of the in situ measurements and the grid scale of the reanalysis products. The correlation and anomaly correlation measure relative variations in the time series including and excluding the seasonal cycle, respectively,

TABLE 5. Soil moisture metrics by individual networks.

		Surface soil moisture					Root zone soil moisture				
		MERRA-2	MERRA-Land	MERRA	ERA-Interim/Land	95% confidence interval (1/2 width)	MERRA-2	MERRA-Land	MERRA	ERA-Interim/Land	95% confidence interval (1/2 width)
ubRMSE ($m^3 m^{-3}$)	SCAN	0.051	0.052	0.058	0.058	0.002	0.036	0.036	0.041	0.039	0.003
	USCRN	0.053	0.054	0.058	0.055	0.013	0.040	0.038	0.044	0.036	0.016
	SMOSMANIA	0.050	0.051	0.055	0.048	0.002	0.045	0.045	0.048	0.042	0.003
	OZNet	0.055	0.055	0.058	0.056	0.007	0.048	0.047	0.056	0.049	0.009
	All sites	0.053	0.053	0.057	0.056	0.002	0.040	0.038	0.043	0.038	0.002
R	SCAN	0.63	0.61	0.52	0.63	0.02	0.65	0.66	0.54	0.62	0.05
	USCRN	0.71	0.71	0.63	0.70	0.15	0.72	0.74	0.64	0.73	0.21
	SMOSMANIA	0.77	0.78	0.74	0.81	0.02	0.75	0.78	0.74	0.77	0.04
	OZNet	0.66	0.66	0.62	0.74	0.08	0.73	0.73	0.62	0.67	0.16
	All sites	0.68	0.68	0.61	0.68	0.02	0.71	0.73	0.63	0.70	0.04
Anomaly *R*	SCAN	0.59	0.57	0.46	0.58	0.02	0.56	0.58	0.43	0.49	0.04
	USCRN	0.64	0.64	0.54	0.64	0.10	0.66	0.69	0.55	0.64	0.15
	SMOSMANIA	0.62	0.62	0.53	0.68	0.02	0.55	0.60	0.52	0.57	0.04
	OZNet	0.61	0.60	0.53	0.68	0.07	0.65	0.65	0.48	0.61	0.12
	All sites	0.62	0.61	0.51	0.62	0.02	0.64	0.66	0.52	0.59	0.03
Bias ($m^3 m^{-3}$)	SCAN	0.058	0.071	0.052	0.086	0.002	0.018	0.035	0.019	0.057	0.003
	USCRN	0.051	0.057	0.033	0.063	0.018	0.010	0.028	0.008	0.039	0.021
	SMOSMANIA	0.034	0.046	0.018	0.098	0.003	0.047	0.069	0.045	0.103	0.003
	OZNet	0.071	0.073	0.038	0.104	0.010	0.048	0.054	0.035	0.032	0.012
	All sites	0.053	0.059	0.036	0.068	0.002	0.016	0.031	0.013	0.048	0.003

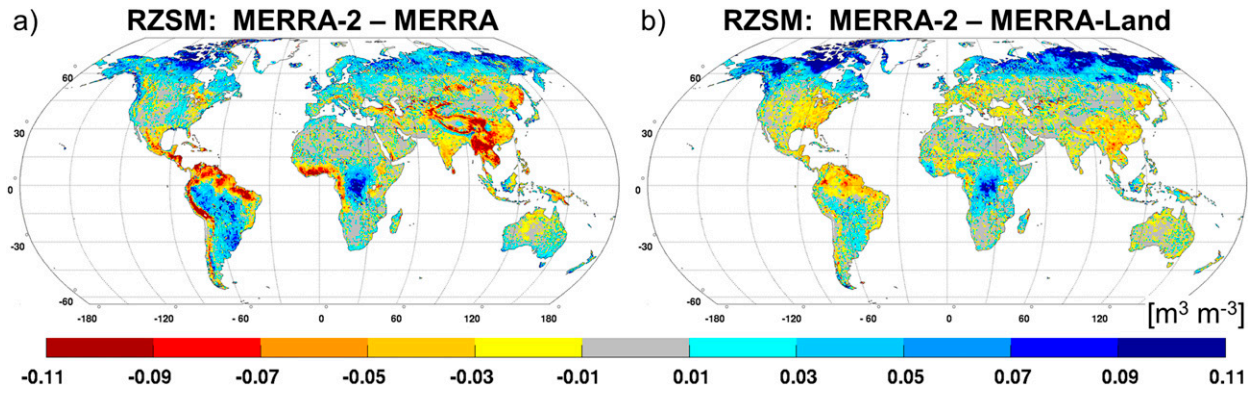

FIG. 3. Annual average root zone soil moisture ($m^3\,m^{-3}$) differences (1980–2015): (a) MERRA-2 minus MERRA and (b) MERRA-2 minus MERRA-Land.

independent of the time-mean amplitude of the variations. Therefore, the improvements brought about by the precipitation corrections are greater in terms of the correlation and anomaly correlation skill metrics (Figs. 2b,c). Here, the MERRA-2 and MERRA-Land surface and root zone soil moisture estimates are significantly better than the MERRA estimates, with increases of ~0.07 for R and ~0.1 for anomaly R. Across all three metrics, the MERRA-2 and MERRA-Land surface soil moisture estimates have nearly identical skill, while MERRA-2 root zone soil moisture estimates are only slightly (not significantly) worse than MERRA-Land estimates.

MERRA-2 and MERRA-Land have nearly identical precipitation forcing at the soil moisture measurement sites, which is the key reason for their generally similar soil moisture performance. The small (nonsignificant) degradation in MERRA-2 root zone soil moisture compared to MERRA-Land is most likely due to the land model parameter changes (Table 2). A similar small degradation was observed by De Lannoy et al. (2014, their Table 4) in their (land only) BLM2 experiment compared to their ''BL'' experiment; these two otherwise identical experiments use the MERRA-2 and MERRA-Land soil parameters, respectively. Note that the small degradation occurs despite the generally improved MERRA-2 surface radiation (Bosilovich et al. 2015) and the more consistent land surface forcing data in the MERRA-2 system (Reichle et al. 2017). Nevertheless, the results confirm that the soil moisture estimates from MERRA-2 are an adequate replacement for those from the land-only MERRA-Land product.

The skill of the soil moisture estimates from ERA-Interim/Land generally lies between that of MERRA and MERRA-2. There are, of course, many differences between the MERRA and ERA-Interim/Land systems, including the modeling and analysis components and the spatial resolution [section 2b(6)]. One important difference is the fact that the precipitation corrections in ERA-Interim/Land are applied to monthly totals, whereas they are applied to daily totals in MERRA-2 and MERRA-Land (except for Africa, where corrections are based on pentads, but Fig. 2 does not include stations in Africa). The precipitation estimates of ERA-Interim/Land therefore depend on the skill of submonthly variations in the precipitation forcing from ERA-I. Since our metrics are based on daily data, it is therefore not surprising that the ERA-Interim/Land estimates are somewhat less skillful than those of MERRA-2 and MERRA-Land, as can be seen, for example, in the ubRMSE values for surface soil moisture (Fig. 2a) and in the anomaly correlations for root zone soil moisture (Fig. 2c).

The relative soil moisture performance of the MERRA products and ERA-Interim/Land is mostly unchanged when individual networks are considered (Table 5). One exception, though, is that the R and anomaly R values for surface soil moisture from ERA-Interim/Land are better, by 0.04 to 0.08, than those of MERRA-2 for the OZNet and SMOSMANIA networks. This is most likely due to the better skill in those regions of the GPCPv2.1 and GPCPv2.2 precipitation products used in ERA-Interim/Land compared to the CPCU product used in MERRA-2. Finally, Table 5 also shows that MERRA and MERRA-2 exhibit the overall lowest soil moisture bias. Specifically, the surface soil moisture bias across all sites is $0.036\,m^3\,m^{-3}$ for MERRA, $0.053\,m^3\,m^{-3}$ for MERRA-2, $0.059\,m^3\,m^{-3}$ for MERRA-Land, and $0.068\,m^3\,m^{-3}$ for ERA-Interim/Land. Similarly, the root zone soil moisture bias across all sites is $0.013\,m^3\,m^{-3}$ for MERRA, $0.016\,m^3\,m^{-3}$ for MERRA-2, $0.031\,m^3\,m^{-3}$ for MERRA-Land, and $0.048\,m^3\,m^{-3}$ for ERA-Interim/Land.

Finally, we briefly address the implications of the changes in the precipitation forcing [section 2a(1); Table 1] and model parameters [section 2a(2); Table 2] on the soil moisture climate in the MERRA products, which may

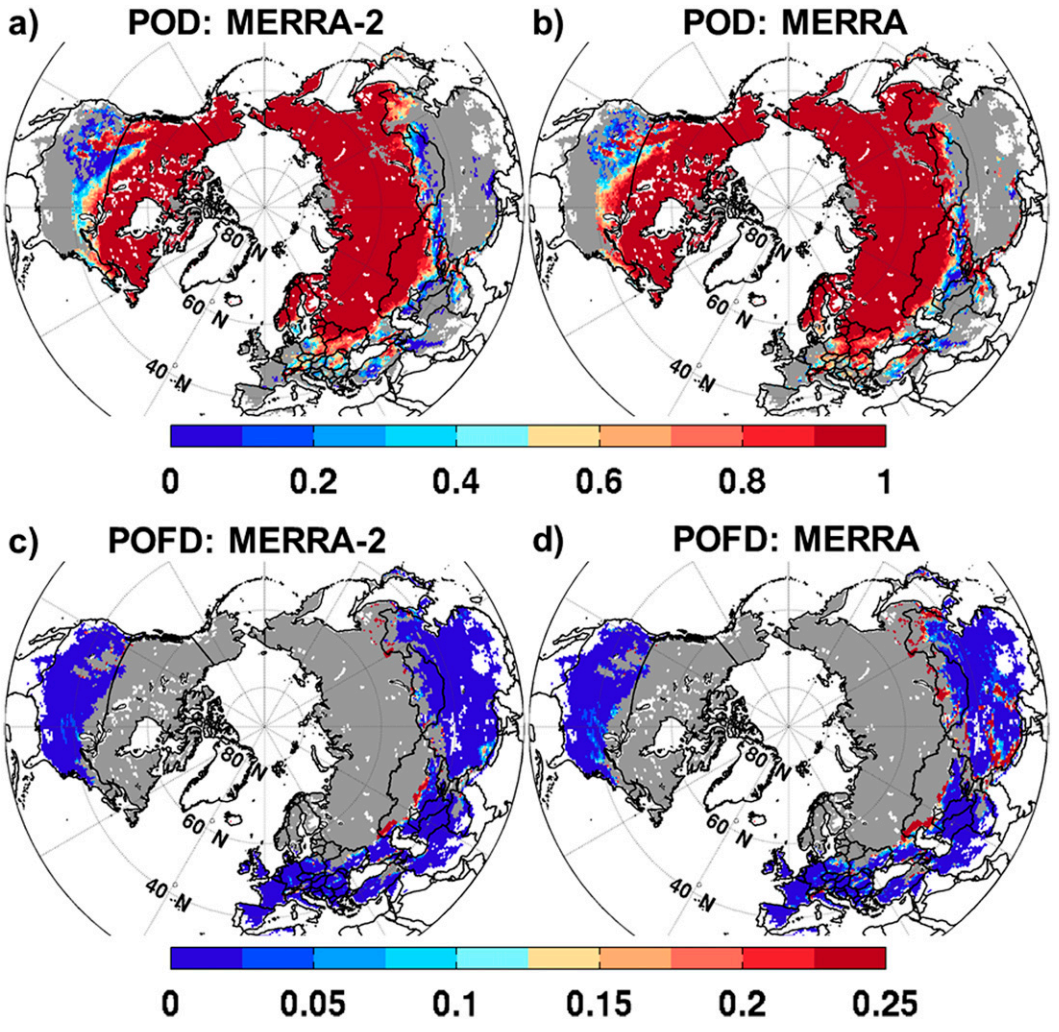

FIG. 4. Snow cover (a),(b) POD and (c),(d) POFD for February based on 2001–15 MODIS observations for (a),(c) MERRA-2 and (b),(d) MERRA. Gray shading indicates areas that had fewer than 30 daily observations in the 15-yr period contributing to the POD or POFD metric. White shading indicates ocean, permanent ice, and land areas that during the entire 15-yr period (1 Oct 2000–31 May 2015) had fewer than 30 daily observations or were never fully snow covered in the MODIS observations or any of the reanalysis datasets.

have important implications for applications that use these data. Figure 3a shows the long-term mean differences in root zone soil moisture between MERRA-2 and MERRA. MERRA-2 soil moisture is considerably wetter than MERRA at high latitudes and in the central Amazon and Congo basins. MERRA-2 is much drier than MERRA in Central America, along the coasts of tropical South America and Africa, and in southeastern Asia. These soil moisture differences are primarily driven by the strong differences in the precipitation forcing between MERRA-2 and MERRA (see Fig. 3d of Reichle et al. 2017).

Next, Fig. 3b shows the root zone soil moisture differences between MERRA-2 and MERRA-Land, which at high latitudes and in Africa are also driven by the differences in precipitation forcing (see Fig. 3c of Reichle et al. 2017). However, MERRA-2 and MERRA-Land also exhibit modest soil moisture differences in areas where their precipitation forcing is the same (i.e., in the low and midlatitudes, except Africa). In most of this area, including the eastern United States, northern South America, and China, MERRA-2 is drier than MERRA-Land by around $0.05\,m^3\,m^{-3}$. The most likely reasons for the drier conditions in MERRA-2 are the changes in the vertical decay factor for the saturated hydraulic conductivity and in the surface turbulence parameterization (Table 2), which act to enhance soil drainage and drying.

c. *Snow*

Snow estimates deserve particular attention because of the changes in high-latitude precipitation and snow

model parameters in the MERRA systems. Our assessment uses MODIS snow cover observations and the CMC snow analysis product, which is based on measurements where available, including in situ snow depth observations [section 2b(3)]. The MODIS data reveal whether the reanalysis snow is placed at the right time and in the right location based on a purely observational reference. The CMC data permit an assessment of the snow amounts against a widely accepted (but not purely observational) data product. It should be noted again that none of the reanalysis products evaluated here includes a snow analysis.

First, Fig. 4 shows the POD and POFD versus MODIS snow cover observations [section 2b(3)] for MERRA-2 and MERRA during the month of February based on 2001–15 daily data. Both products generally exhibit high POD values of 0.9 and greater at latitudes north of about 50°N (Figs. 4a,b). POD values are much lower and typically less than 0.5 along and below 45°N in the interior of Asia and in western North America, including the Canadian prairies. Overall, MERRA-2 has a lower (worse) average POD (0.86) than MERRA (0.91). The POFD values for February, shown in Figs. 4c,d are typically below 0.025, but in Asia along 50°N there are isolated patches of POFD values as high as and exceeding 0.25. Overall, MERRA-2 has substantially fewer patches of high POFD and a lower (better) average POFD value (0.012) than MERRA (0.035). The February POD and POFD metrics for MERRA-Land are very similar to those for MERRA-2 (not shown).

Next, Fig. 5a shows time series of monthly climatologies for snow-covered area from MODIS and the MERRA products. Throughout the year, all MERRA products slightly underestimate the snow-covered area (Fig. 5a) by 10%–30%, with MERRA-Land generally having the largest (negative) bias, MERRA the smallest, and MERRA-2 falling between the two but generally closer to MERRA-Land. The snow cover bias is also reflected in the time series of POD and POFD versus MODIS (Figs. 5b,c), with MERRA-2 typically having lower (worse) POD values but also lower (better) POFD values than MERRA. Exceptions occur in April and May, when the snow cover area bias is nearly the same for all MERRA products, but the MERRA-2 POD values are higher than those of MERRA and MERRA-Land. Moreover, during the early season (October and November), MERRA-2 snow cover area and POD values are noticeably better than those of MERRA-Land, which, again, has a more pronounced negative bias.

The differences in snow cover performance between MERRA-2, MERRA-Land, and MERRA are due to the differences in the forcing, primarily precipitation and in the snow model parameters, primarily the change

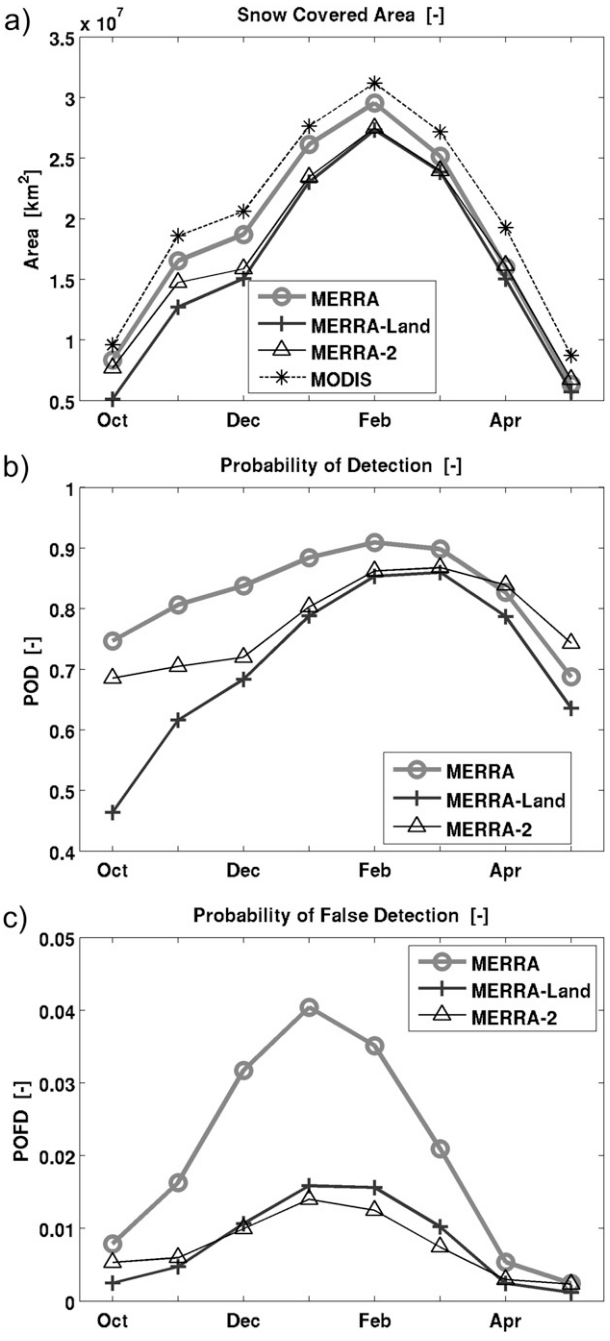

FIG. 5. (a) Mean total land area that is covered by snow, excluding glaciated surfaces, lake ice, and areas not observed by MODIS (e.g., during polar night). (b) Mean snow cover POD vs MODIS. (c) Mean snow cover POFD vs MODIS. Mean values are computed using data from 1 Oct 2000 to 31 May 2015.

from the WEMIN value of $13\,\mathrm{kg\,m^{-2}}$ in MERRA to $26\,\mathrm{kg\,m^{-2}}$ in MERRA-Land and MERRA-2 [Table 2; section 2a(2)]. Because of their higher WEMIN value, MERRA-2 and MERRA-Land tend to have lower SCF values than MERRA under low snow conditions. For

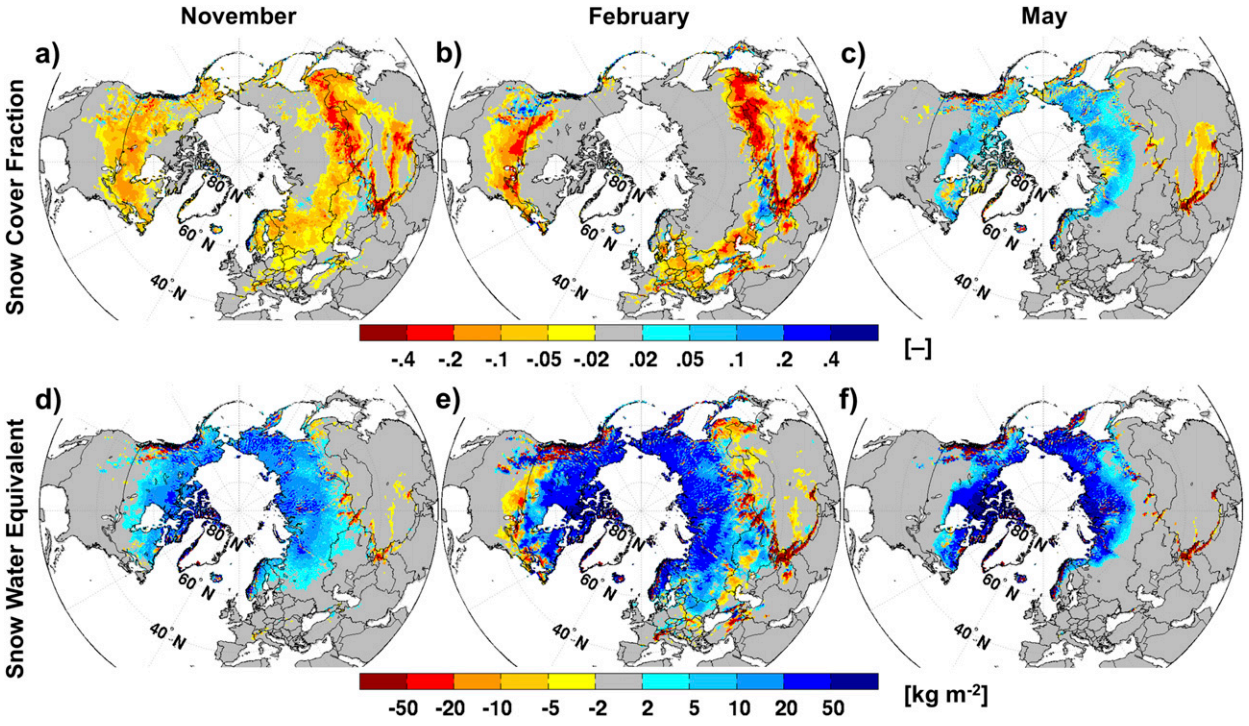

FIG. 6. (a) Mean difference between MERRA-2 and MERRA SCF for November. (b) As in (a), but for February. (c) As in (a), but for May. (d) As in (a), but for SWE. (e) As in (d), but for February. (f) As in (d), but for May. Mean values are computed using data from 1 Oct 2000 to 31 May 2015.

example, in November MERRA-2 has much lower SCF than MERRA (Fig. 6a) nearly everywhere, despite MERRA-2 having greater SWE than MERRA (Fig. 6d). In the middle of the snow season (February), the much lower SCF in MERRA-2 (compared to MERRA) along and below 45°N in the interior of Asia and in western North America, including the Canadian prairies (Fig. 6b), corresponds to lower SWE in MERRA-2 (Fig. 6e), which exacerbates the impact of the higher WEMIN parameter in MERRA-2. Late in the snow season, by contrast, the May SCF for MERRA-2 exceeds that for MERRA at high latitudes (Fig. 6c) because MERRA-2 has much more SWE than MERRA in that region (Fig. 6f)—so much more, in fact, as to overcome the impact of the higher WEMIN parameter in MERRA-2.

Next, Fig. 7 provides an assessment of SWE skill for MERRA-2, MERRA-Land, MERRA, and ERA-Interim/Land against CMC data [section 2b(3)]. For the correct interpretation of this figure, it is important to keep in mind where the CMC snow analysis is informed by in situ snow depth measurements. This is shown in Fig. 8, which indicates for each grid cell the mean distance d to the nearest measurement used in the CMC snow analysis. Most of the in situ measurements are in the midlatitudes, with relatively few measurements available poleward of 60°N except in Scandinavia and southern Alaska. No measurements are available in China. We consider the CMC data reliable and compute the spatially averaged metrics shown in Table 6 only for grid cells with $d \leq 100$ km, shown in dark blue colors in Fig. 8. In these regions, the SWE bias of MERRA-2 versus CMC observations is generally small (Fig. 7a) and only 0.4 kg m^{-2} on average (Table 6). In the high latitudes, MERRA-2 has a pronounced positive bias, with mean SWE values exceeding those of the CMC product by up to 70 kg m^{-2} (Fig. 7a). While the CMC estimates may not be reliable in the high latitudes, this result provides further evidence that the systematic SCF underestimation in MERRA-2 associated with its increased WEMIN value is partially compensated by excessive snow amounts.

The SWE bias for MERRA-Land (Fig. 7c) and MERRA (Fig. 7e) has a spatial pattern similar to that of MERRA-2 but with a slightly stronger negative bias in midlatitudes and a smaller positive bias at high latitudes. On average across the regions where CMC data are reliable, the SWE bias is -9.7 kg m^{-2} for MERRA-Land and -3.4 kg m^{-2} for MERRA (Table 6). ERA-Interim/Land overestimates SWE in most of Eurasia and in northwestern North America and underestimates SWE in most of the rest of North America (Fig. 7g), with

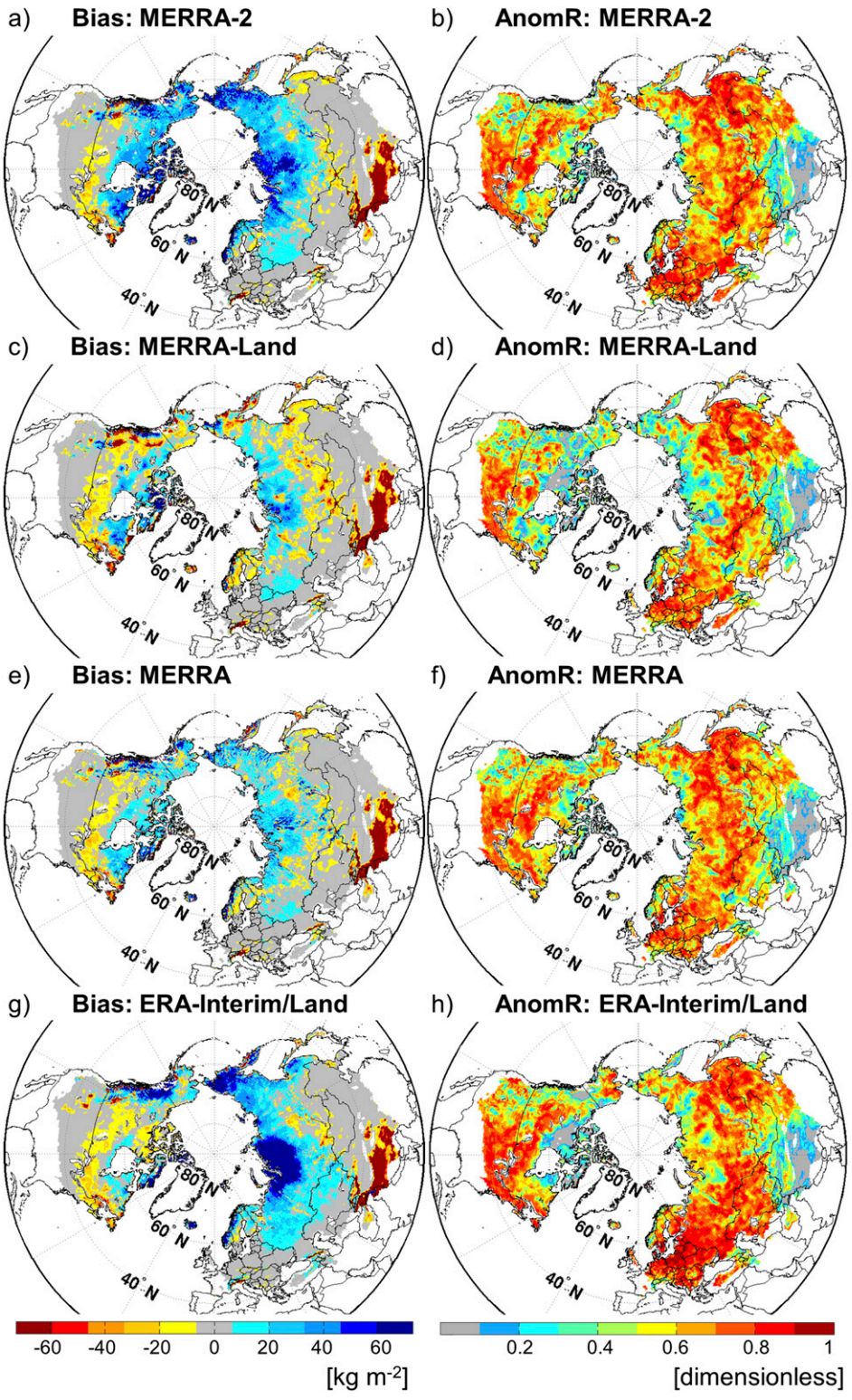

FIG. 7. (a) Bias and (b) anomaly time series correlation coefficient of MERRA-2 SWE estimates vs CMC data [section 2b(3)]. (c),(d) As in (a),(b), but for MERRA-Land. (e),(f) As in (a),(b), but for MERRA. (g),(h) As in (a),(b), but for ERA-Interim/Land. Metrics are computed for August 1998–December 2014.

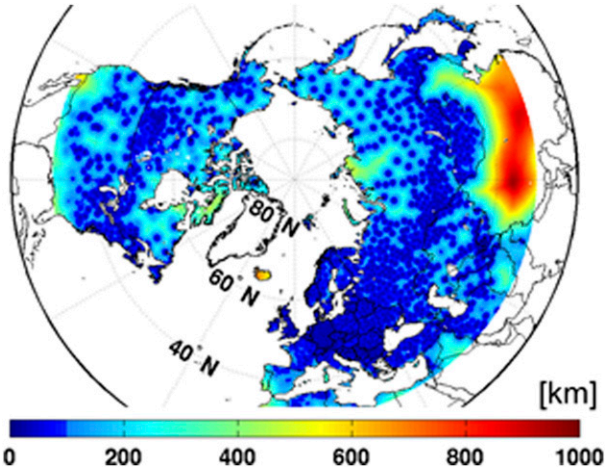

FIG. 8. Mean (2000–09) distance d (km) to the closest in situ measurement used in the CMC snow depth analysis. The metrics of Table 6 are averages only over grid cells with $d \leq 100$ km (shown here in dark blue colors).

an average positive bias of 14.1 kg m^{-2} in the station-covered area (Table 6). Compared to MERRA-2, ERA-Interim/Land has an even stronger positive SWE bias in northcentral Siberia. All reanalysis products share a very strong negative SWE bias against CMC over the Tibetan Plateau. Given the absence of in situ measurements in Tibet (Fig. 8), CMC is likely to overestimate snow amounts in that region (R. Brown 2016, personal communication).

Figure 7 also shows SWE skill in terms of the anomaly R values versus daily CMC estimates, reflecting the ability of the reanalysis products to estimate synoptic events and interannual variability. The MERRA-2 anomaly R values range from 0.5 to 0.8 in most regions except Tibet, where skill values are close to zero (Fig. 7b). The highest skill values are in eastern North America and most of Europe, where the precipitation gauge network is relatively dense. These areas of good skill (for which the average anomaly R value is 0.66; Table 6) also coincide with the regions that have the highest density of in situ snow depth measurements supporting the CMC estimates (Fig. 8).

The anomaly R value for MERRA-Land SWE is 0.61 on average (in areas within 100 km of in situ measurements; Table 6) and thus lower than that of MERRA-2. The MERRA-Land SWE skill is lower than that of MERRA-2 primarily in eastern Siberia, northern and western Canada, and Alaska (Fig. 7d). In these regions, the CPCU product that was used to correct the MERRA-Land precipitation is based on very few gauges and subject to large errors. The improved MERRA-2 SWE skill compared to MERRA-Land thus further supports the decision not to correct the precipitation in MERRA-2 at high latitudes. [Note that the climatological discontinuities in MERRA-Land SWE in 1982 and in 1998 shown in Fig. 2 of Mudryk et al. (2015) are not captured in the evaluation against CMC data, which start in 1998.] The MERRA SWE skill is 0.62 on average in the areas within 100 km of in situ snow measurements (Table 6). It is similar to that of MERRA-2 except for the Canadian Archipelago, where MERRA is less skillful than MERRA-2 (Fig. 7f). Finally, the skill of the ERA-Interim/Land SWE estimates is 0.67 on average (Table 6), which is slightly higher than that of MERRA-2. ERA-Interim/Land has higher skill than MERRA-2 in western Eurasia and eastern North America but lower skill than MERRA-2 in central Eurasia and in the few high-latitude areas that have snow measurements (Fig. 7h).

Table 6 provides additional skill metrics versus CMC SWE data, including ubRMSE and R values, as well as metrics for snow depth. The relative performance of the reanalysis snow depth products is very similar to that of the SWE estimates (which suggests that the conversion of the CMC snow depth data into SWE using climatological snow densities is indeed acceptable). ERA-Interim/Land has the best R value of the four products, while MERRA-2 and MERRA are best in terms of ubRMSE. Across all metrics, MERRA-Land tends to do the worst, which provides further evidence that the CPCU precipitation product should not be used at high latitudes.

d. Streamflow

While the MERRA data products generally underestimate runoff and overestimate latent heat fluxes

TABLE 6. Skill metrics vs CMC data. Metrics are area-weighted spatial averages over grid cells for which nearby in situ snow depth measurements contribute to the CMC analysis. See Fig. 8 for coverage.

	Bias		ubRMSE		R		Anomaly R	
	SWE	Snow depth	SWE	Snow depth	SWE	Snow depth	SWE	Snow depth
	(kg m^{-2})	(m)	(kg m^{-2})	(m)	(dimensionless)			
MERRA-2	0.4	0.007	29.2	0.081	0.82	0.82	0.66	0.67
MERRA-Land	−9.7	−0.021	29.9	0.083	0.79	0.80	0.61	0.62
MERRA	−3.4	−0.004	29.0	0.080	0.79	0.80	0.62	0.64
ERA-Interim/Land	14.1	0.106	31.6	0.094	0.82	0.84	0.67	0.69

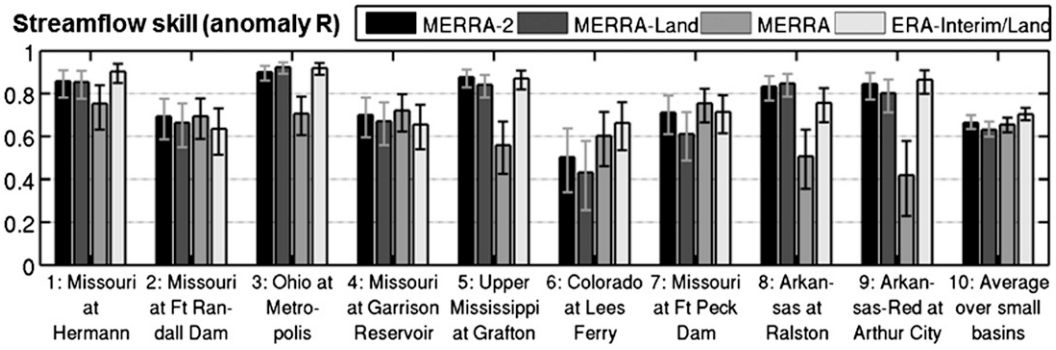

FIG. 9. Seasonal anomaly time series correlation coefficients (dimensionless) for runoff estimates from MERRA-2, MERRA-Land, MERRA, and ERA-Interim/Land. See Table 4 for details about the basins, which are in order of decreasing size.

(Lorenz and Kunstmann 2012; Reichle et al. 2017), they nevertheless provide value in terms of the estimated time series variations. Figure 9 shows the streamflow skill for MERRA-2, MERRA-Land, MERRA, and ERA-Interim/Land estimates, measured as the anomaly time series correlation coefficient against naturalized streamflow observations [section 2b(4)]. The figure shows skill for basins ordered by decreasing size.

Generally, the MERRA-2 streamflow skill (anomaly R) values range from 0.5 to 0.9 (Fig. 9). In four of the nine major basins assessed here (Ohio, upper Mississippi, Arkansas, and Arkansas-Red), the MERRA-2 skill significantly exceeds that of MERRA, owing to both MERRA-2 precipitation corrections and model improvements. On balance, the MERRA-2 streamflow skill also slightly exceeds that of MERRA-Land, albeit not with statistical significance. Since the precipitation forcing in MERRA-2 and MERRA-Land is essentially the same for the basins examined here (except for the tapering back to the AGCM-generated precipitation in

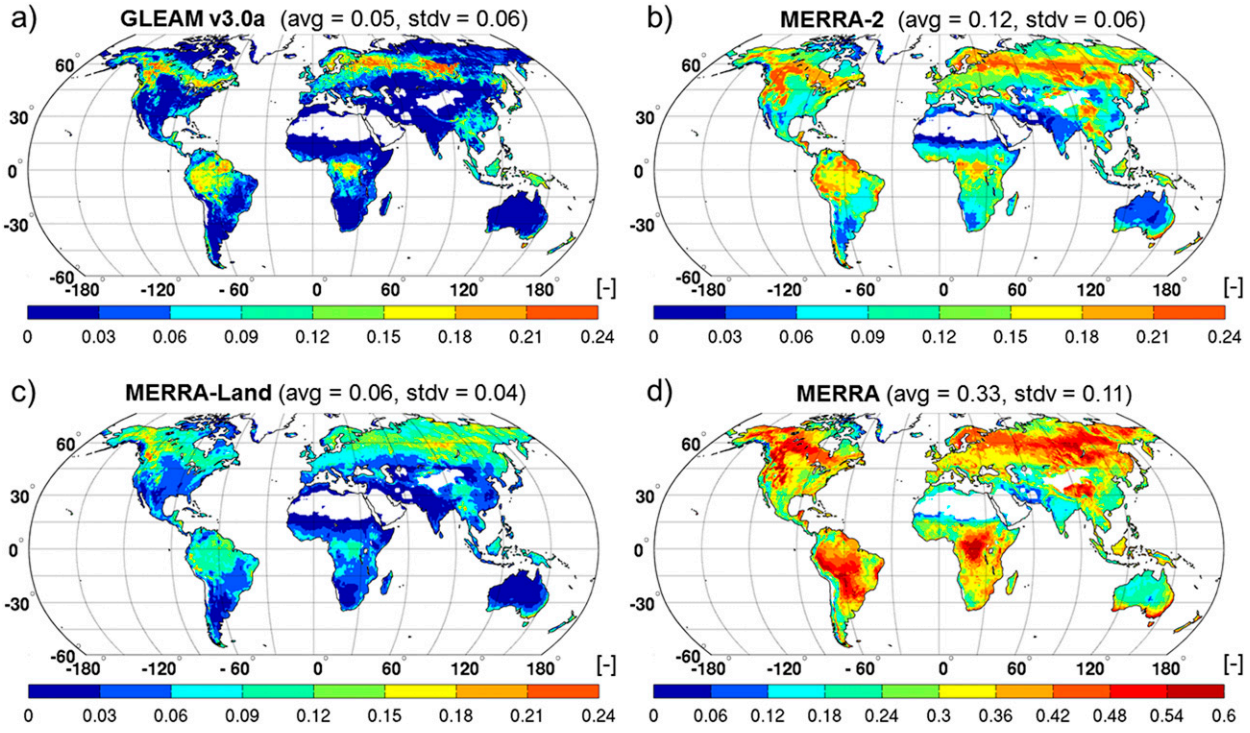

FIG. 10. Annual mean (1980–2014) interception loss fraction (dimensionless) from (a) GLEAM v3.0a, (b) MERRA-2, (c) MERRA-Land, and (d) MERRA. Titles indicate the spatial average and standard deviation across each map. Regions with annual mean MERRA-2 rainfall less than 0.2 mm day^{-1} were masked out. Note the different color scale for MERRA.

MERRA-2 poleward from 42.5° latitude; Reichle et al. 2017), the small improvements in skill from MERRA-Land to MERRA-2 can be attributed primarily to the changes in the MERRA-2 Catchment model parameters [Table 2; section 2a(2); De Lannoy et al. 2014]. Finally, the average streamflow skill of ERA-Interim/Land is almost identical to that of MERRA-2.

e. Interception loss fraction

As discussed by Reichle et al. (2011), one major motivation for the development of MERRA-Land was the significant overestimation in MERRA of the interception loss fraction I, defined as the fraction of incoming rainfall intercepted by the canopy and reevaporated [section 2b(5)]. The estimated I values are very sensitive to the treatment of canopy interception in the land surface model and to the rainfall and radiation forcing. It is thus useful to assess the interception loss fraction in MERRA-2. As a reference, Fig. 10a shows climatological estimates of I from GLEAM data [section 2b(5)]. The largest I values are found in the boreal forests of North America, Scandinavia, and Russia, ranging from $I = 0.12$ to $I = 0.24$. Somewhat smaller values are found in the southeastern United States and parts of Europe as well as in the tropical rain forests of Southeast Asia and the Amazon and Congo basins, with values ranging from $I = 0.06$ to $I = 0.21$. The corresponding map for MERRA is shown in Fig. 10d (note the different color scale). As summarized in Reichle et al. (2011), MERRA's average I values are greater than 0.24 almost everywhere, even in nonforested areas (e.g., the U.S. Great Plains), and even exceed 0.5 in some regions. Globally averaged, MERRA's interception loss fraction is $I = 0.33$. While there is considerable uncertainty in global interception products, including GLEAM estimates, the MERRA I values are obviously too high and not realistic.

The I values of MERRA-2 (Fig. 10b) are far more realistic than those of MERRA. The global average for MERRA-2 is $I = 0.12$, which is higher than the global average $I = 0.05$ suggested by GLEAM but matches the value of $I = 0.12$ reported by Sakaguchi and Zeng (2009) for the Community Land Model, version 3.5. The spatial pattern of I from MERRA-2 is similar to that of GLEAM, with the highest values in forested areas. The generally higher interception loss in MERRA-2 is partly the result of the subdaily (or subpentad) temporal distribution of the rainfall forcing. Even though the daily (or pentad) precipitation totals in MERRA-2 are prescribed from observations [section 2a(1)], the intensity of rainfall in MERRA-2 is too low at subdaily (or subpentad) time scales, particularly for convective rainfall events (Reichle et al. 2017). This underestimation of rainfall intensity in MERRA-2 is due to the use of the MERRA diurnal precipitation as the background field in the precipitation corrections algorithm and contributes to the overestimation of the I values in MERRA-2. Moreover, Martens et al. (2016, manuscript submitted to *Geosci. Model. Dev. Discuss.*) report that GLEAM v3.0a has less interception than earlier GLEAM versions because its updated, higher-resolution land-cover inputs have larger bare-soil fractions than earlier GLEAM products and, presumably, MERRA-2.

Even though the canopy interception model is the same in MERRA-2 and MERRA-Land, there are noticeable differences between the interception loss fractions in the two systems. While the global average value of $I = 0.06$ for MERRA-Land is close to that of GLEAM ($I = 0.05$), the spatial distribution of the interception loss fraction in MERRA-Land appears too smooth, with some overestimation of the lower values in nonforested midlatitude regions and an underestimation of the higher values in the boreal and tropical forests (Fig. 10c). The differences in I between MERRA-Land and MERRA-2 stem from at least three factors. First, the difference in the surface turbulence schemes of MERRA-Land (Louis 1979) and MERRA-2 [Helfand and Schubert 1995; section 2a(2)] impacts the reevaporation of the water intercepted by the canopy. Second, the surface radiation forcing is generally more realistic in MERRA-2 than in MERRA-Land (Reichle et al. 2017). Third, the differences in I at high latitudes and in tropical Africa are partly a consequence of the differences in the precipitation forcing there (see Fig. 3c of Reichle et al. 2017).

4. Summary and conclusions

The recent MERRA-2 atmospheric reanalysis provides global, 1-hourly estimates of land surface conditions for 1980 to present at about 50-km spatial resolution. MERRA-2 replaces the MERRA data product, which was first published in 2010. In this paper we investigated the skill of the MERRA-2 land surface hydrology estimates, including terrestrial water storage, soil moisture, snow, runoff, and rainfall interception. The MERRA-2 land surface is forced with precipitation that is a merger of observational data products and the precipitation generated by the AGCM within the MERRA-2 system. Because of the precipitation corrections and because of improvements in the land surface model, the skill of the MERRA-2 land surface hydrology estimates (vs independent data) is generally (but not always) greater than that of MERRA estimates. MERRA-2 also shows improvements compared to MERRA-Land, a land-only reanalysis product that supplemented MERRA and benefited from earlier

versions of the precipitation corrections and land model improvements.

Monthly terrestrial water storage (TWS) estimates from MERRA-2 show better R and anomaly R values versus GRACE observations than do those from MERRA (Fig. 1). The improvements in MERRA-2 are most pronounced for South America and Africa, where the (corrected) MERRA-2 precipitation is most improved compared to MERRA precipitation. MERRA-2 TWS estimates are also better than those of MERRA-Land, with improvements driven mainly by the improved MERRA-2 precipitation corrections over Africa and the (intentional) lack of (gauge based) corrections at high latitudes. However, the evaluation against GRACE observations also reveals how poor precipitation gauge data over Myanmar adversely impact the skill of MERRA-2 TWS estimates. Furthermore, daily surface and root zone soil moisture estimates from MERRA-2 have good skill against in situ measurements from between 220 and 320 stations in the United States, France, and Australia (Fig. 2; Table 5). Generally, the MERRA-2 soil moisture skill (in terms of ubRMSE, R, and anomaly R values) is significantly better than that of MERRA, somewhat better than that of ERA-Interim/Land, and similar to that of MERRA-Land. Because of the changes in precipitation forcing and soil parameters, the soil moisture climatologies of MERRA-2, MERRA-Land, and MERRA differ (Fig. 3), which is important to consider when the data are used in applications.

The evaluation of MERRA-2 daily snow cover skill (against MODIS observations) reveals that MERRA-2 has both a lower (worse) POD and a lower (better) POFD than MERRA (Figs. 4 and 5). This result is a combination of two factors. First, in the middle of the snow season, MERRA-2 has less snow water equivalent (SWE) than MERRA at the lower latitudes (Fig. 6e), where snow cover is most dynamic throughout the cold season and thus most difficult to estimate. (MERRA-2 SWE exceeds that of MERRA in the global average.) Second, the WEMIN snow model parameter, which governs the model's snow accumulation and depletion curve, is twice as large in MERRA-2 as in MERRA, which contributes to the systematic underestimation of the snow cover fraction in MERRA-2, particularly early in the snow season. POD values for MERRA-2 are better than those of MERRA only late in the season, when the excessive MERRA-2 SWE accumulation at high latitudes takes longer to melt. A comparison of the mean MERRA-2 SWE and snow depth against the independent CMC product also suggests that MERRA-2 has too much snow at high latitudes and not quite enough at midlatitudes (Fig. 7; Table 6). Nevertheless, in terms of SWE and snow depth metrics against CMC data, the MERRA-2 skill is somewhat greater than that of MERRA-Land and MERRA and comparable to that of ERA-Interim/Land.

Seasonally averaged runoff estimates from MERRA-2 show good skill (in terms of anomaly time series correlations) against naturalized streamflow estimates from nine large and nine small basins in the United States (Fig. 9; Table 4). The skill of MERRA-2 is significantly better than that of MERRA at four of the nine large basins. On average, MERRA-2 runoff skill is somewhat better than that of MERRA-Land and is comparable to that of ERA-Interim/Land. Finally, the interception loss fraction estimates of MERRA-2 exhibit good skill against an independent data product, with MERRA-2 being much better than MERRA and comparable to MERRA-Land (Fig. 10).

Preliminary results from a forthcoming publication suggest that MERRA-2 has smaller biases in land surface turbulent heat fluxes than MERRA compared to multiple independent reference datasets. The MERRA-2 latent heat fluxes also have consistently improved monthly anomaly correlations versus the same independent datasets, including in regions where evapotranspiration is moisture limited (and thus sensitive to improved precipitation) and regions where it is not (reflecting the importance of model parameter changes, specifically those associated with rainfall interception). It is thus particularly encouraging to note that the enhancements in the MERRA-2 land surface hydrology estimates are paired with better estimates of the land surface energy balance.

Looking ahead, there are several avenues for further improvements in land surface hydrology estimates from reanalysis systems, including improvements in the approach used to impose precipitation corrections. Outside of the high latitudes, the corrected MERRA-2 land surface precipitation forcing relies on high-quality, daily (or pentad) observational data products that go back to the 1980s and are updated to the present with latency on the order of weeks. In the high latitudes, however, such data products are not available, and, unfortunately, the MERRA-2 AGCM-generated precipitation exhibits a pronounced wet bias in spring and summer (Reichle et al. 2017). One possible way to improve high-latitude precipitation is to apply climatological rescaling factors to the AGCM-generated precipitation based on long-term, monthly data products such as those from the Global Precipitation Climatology Project (Huffman et al. 2009). This approach was not an option for MERRA-2 because it requires knowledge of the climate of the reanalysis precipitation prior to its generation. Climatological precipitation rescaling in the high latitudes, however, is an option for land-only data products such as

the SMAP level-4 soil moisture product (Reichle et al. 2016) or, possibly, a supplemental, land-only "MERRA-2/Land" reanalysis (similar to what MERRA-Land is for MERRA). Users of MERRA-2 atmospheric forcing with a particular interest in the high latitudes may likewise benefit from climatological rescaling of the precipitation. The approach might also work for reanalysis systems that have been used in Numerical Weather Prediction operations for several years prior to the generation of the reanalysis.

Not counting the precipitation corrections, MERRA-2 does not include a land surface analysis. ERA-I (but not ERA-Interim/Land) includes a simple snow analysis and an analysis of screen-level air temperature and humidity observations. Land data assimilation has matured to the point where it should be possible to include more comprehensive and sophisticated land surface data assimilation elements in future atmospheric (or coupled atmosphere–ocean) reanalysis systems. The assimilation of soil moisture and snow observations in particular complements the use of precipitation observations. This complementarity will be most valuable where the precipitation gauge network is sparse or when precipitation occurs as snowfall, which is difficult to measure. Future global reanalysis systems are also likely to include estimates of land carbon (and possibly nitrogen) fluxes and stores, given that dynamic vegetation phenology modules are now being included in the land surface components of global NWP systems (Boussetta et al. 2013; Koster and Walker 2015).

Acknowledgments. Funding for this work was provided by the NASA Modeling, Analysis, and Prediction program. Computational resources were provided by the NASA High-End Computing Program through the NASA Center for Climate Simulation. MERRA data products are disseminated by the NASA Goddard Earth Sciences Data and Information Services Center. We are grateful for access to the many datasets that supported this work and highly appreciate all those who made them possible, including USDA, NOAA National Centers for Environmental Information (formerly NCDC), NASA GSFC, NOAA CPC, International Soil Moisture Network, Environment Canada CMC, U.S. Army Corps of Engineers, U.S. Bureau of Reclamation, Columbia River Basin Climate Change Scenarios Database, California Data Exchange Commission, and ECMWF. GRACE land products are supported by the NASA MEaSUREs program. Special thanks go to R. Brown for valuable assistance with the CMC snow analysis data, to E. Dutra and G. Balsamo for help with the ERA-Interim/Land data, to D. Miralles for help with the GLEAM data, to J. Walker and X. Wu for the Oznet data, and to E. Maurer for the naturalized streamflow data. Thanks also to three anonymous reviewers for valuable comments.

REFERENCES

Albergel, C., and Coauthors, 2008: From near-surface to root-zone soil moisture using an exponential filter: An assessment of the method based on in situ observations and model simulations. *Hydrol. Earth Syst. Sci.*, **12**, 1323–1337, doi:10.5194/hess-12-1323-2008.

Balsamo, G., and Coauthors, 2015: ERA-Interim/Land: A global land surface reanalysis data set. *Hydrol. Earth Syst. Sci.*, **19**, 389–407, doi:10.5194/hess-19-389-2015.

Beck, H. E., and Coauthors, 2017: MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrol. Earth Syst. Sci.*, **21**, 589–615, doi:10.5194/hess-2016-236.

Bell, J., and Coauthors, 2013: U.S. Climate Reference Network soil moisture and temperature observations. *J. Hydrometeor.*, **14**, 977–988, doi:10.1175/JHM-D-12-0146.1.

Bosilovich, M. G., and Coauthors, 2015: MERRA-2: Initial evaluation of the climate. NASA Tech. Rep. NASA/TM-2015-104606, Vol. 43, 139 pp. [Available online at https://gmao.gsfc.nasa.gov/pubs/docs/Bosilovich803.pdf.]

——, R. Lucchesi, and M. Suarez, 2016: MERRA-2: File specification. NASA GMAO Office Note 9, 75 pp. [Available online at https://gmao.gsfc.nasa.gov/pubs/docs/Bosilovich785.pdf.]

Boussetta, S., and Coauthors, 2013: Natural land carbon dioxide exchanges in the ECMWF integrated forecasting system: Implementation and offline validation. *J. Geophys. Res. Atmos.*, **118**, 5923–5946, doi:10.1002/jgrd.50488.

Brasnett, B., 1999: A global analysis of snow depth for numerical weather prediction. *J. Appl. Meteor.*, **38**, 726–740, doi:10.1175/1520-0450(1999)038<0726:AGAOSD>2.0.CO;2.

Brown, R. D., and B. Brasnett, 2010: Canadian Meteorological Centre (CMC) daily snow depth analysis data, version 1. National Snow and Ice Data Center, accessed 1 July 2016, doi:10.5067/W9FOYWH0EQZ3.

Chen, M., W. Shi, P. Xie, V. B. S. Silva, V. E. Kousky, R. Wayne Higgins, and J. E. Janowiak, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. *J. Geophys. Res.*, **113**, D04110, doi:10.1029/2007JD009132.

Cosby, B., G. Hornberger, R. Clapp, and T. Ginn, 1984: A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soil. *Water Resour. Res.*, **20**, 682–690, doi:10.1029/WR020i006p00682.

Cullather, R. I., S. M. J. Nowicki, B. Zhao, and M. J. Suarez, 2014: Evaluation of the surface representation of the Greenland ice sheet in a general circulation model. *J. Climate*, **27**, 4835–4856, doi:10.1175/JCLI-D-13-00635.1.

De Lannoy, G. J. M., and R. H. Reichle, 2016: Global assimilation of multiangle and multipolarization SMOS brightness temperature observations into the GEOS-5 Catchment land surface model for soil moisture estimation. *J. Hydrometeor.*, **17**, 669–691, doi:10.1175/JHM-D-15-0037.1.

——, R. D. Koster, R. H. Reichle, S. P. P. Mahanama, and Q. Liu, 2014: An updated treatment of soil texture and associated hydraulic properties in a global land modeling system. *J. Adv. Model. Earth Syst.*, **6**, 957–979, doi:10.1002/2014MS000330.

Dee, D., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation

system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.

Diamond, H., and Coauthors, 2013: U.S. Climate Reference Network after one decade of operations: Status and assessment. *Bull. Amer. Meteor. Soc.*, **94**, 485–498, doi:10.1175/BAMS-D-12-00170.1.

Dirmeyer, P., and T. Oki, 2002: The Second Global Soil Wetness Project (GSWP-2) science and implementation plan. IGPO Publ. 37, 64 pp.

Dorigo, W. A., and Coauthors, 2011: The International Soil Moisture Network: A data hosting facility for global in situ soil moisture measurements. *Hydrol. Earth Syst. Sci.*, **15**, 1675–1698, doi:10.5194/hess-15-1675-2011.

Entekhabi, D., R. H. Reichle, R. D. Koster, and W. T. Crow, 2010: Performance metrics for soil moisture retrievals and application requirements. *J. Hydrometeor.*, **11**, 832–840, doi:10.1175/2010JHM1223.1.

Gash, J. H. C., 1979: An analytical model of rainfall interception by forests. *Quart. J. Roy. Meteor. Soc.*, **105**, 43–55, doi:10.1002/qj.49710544304.

GMAO, 2015a: MERRA-2 tavgM_2d_lnd_Nx: 2d, Monthly mean, time-averaged, single-level, assimilation, land surface diagnostics, V5.12.4. Goddard Earth Sciences Data and Information Services Center (GES DISC), accessed 1 September 2016, doi:10.5067/8S35XF81C28F.

——, 2015b: MERRA-2 tavg1_2d_lnd_Nx: 2d, 1-hourly, time-averaged, single-level, assimilation, land surface diagnostics, V5.12.4. Goddard Earth Sciences Data and Information Services Center (GES DISC), accessed 1 September 2016, doi:10.5067/RKPHT8KC1Y1T.

——, 2015c: MERRA-2 const_2d_asm_Nx: 2d, constants, V5.12.4. Goddard Earth Sciences Data and Information Services Center (GES DISC), accessed 1 September 2016, doi:10.5067/ME5QX6Q5IGGU.

Hall, D. K., V. V. Salomonson, and G. A. Riggs, 2006: MODIS/*Terra* snow cover daily L3 global 0.05° CMG, version 5. NASA National Snow and Ice Data Center Distributed Active Archive Center, accessed 1 June 2016, doi:10.5067/EI5HGLM2NNHN.

Helfand, H. M., and S. D. Schubert, 1995: Climatology of the simulated Great Plains low-level jet and its contribution to the continental moisture budget of the United States. *J. Climate*, **8**, 784–806, doi:10.1175/1520-0442(1995)008<0784:COTSGP>2.0.CO;2.

Huffman, G. J., R. F. Adler, D. T. Bolvin, and G. Gu, 2009: Improving the global precipitation record: GPCP version 2.1. *Geophys. Res. Lett.*, **36**, L17808, doi:10.1029/2009GL040000.

Koster, R. D., and G. K. Walker, 2015: Interactive vegetation phenology, soil moisture, and monthly temperature forecasts. *J. Hydrometeor.*, **16**, 1456–1465, doi:10.1175/JHM-D-14-0205.1.

——, M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2000: A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure. *J. Geophys. Res.*, **105**, 24 809–24 822, doi:10.1029/2000JD900327.

Landerer, F. W., and S. C. Swenson, 2012: Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resour. Res.*, **48**, W04531, doi:10.1029/2011WR011453.

Livneh, B., and D. P. Lettenmaier, 2012: Multi-criteria parameter estimation for the Unified Land Model. *Hydrol. Earth Syst. Sci.*, **16**, 3029–3048, doi:10.5194/hess-16-3029-2012.

Lorenz, C., and H. Kunstmann, 2012: The hydrological cycle in three state-of-the-art reanalyses: Intercomparison and performance analysis. *J. Hydrometeor.*, **13**, 1397–1420, doi:10.1175/JHM-D-11-088.1.

Louis, J. E., 1979: A parametric model of vertical eddy fluxes in the atmosphere. *Bound.-Layer Meteor.*, **17**, 187–202, doi:10.1007/BF00117978.

Mahanama, S., B. Livneh, R. Koster, D. Lettenmaier, and R. Reichle, 2012: Soil moisture, snow, and seasonal streamflow forecasts in the United States. *J. Hydrometeor.*, **13**, 189–203, doi:10.1175/JHM-D-11-046.1.

——, and Coauthors, 2015: Land boundary conditions for the Goddard Earth Observing System Model version 5 (GEOS-5) climate modeling system: Recent updates and data file descriptions. NASA Tech. Memo. NASA/TM-2015-104606, Vol. 39, 55 pp. [Available online at https://ntrs.nasa.gov/search.jsp?R=20160002967.]

Miralles, D. G., J. H. Gash, T. R. H. Holmes, R. A. M. de Jeu, and A. J. Dolman, 2010: Global canopy interception from satellite observations. *J. Geophys. Res.*, **115**, D16122, doi:10.1029/2009JD013530.

——, T. R. H. Holmes, R. A. M. de Jeu, J. H. Gash, A. G. C. A. Meesters, and A. J. Dolman, 2011: Global land-surface evaporation estimated from satellite-based observations. *Hydrol. Earth Syst. Sci.*, **15**, 453–469, doi:10.5194/hess-15-453-2011.

Molod, A., L. Takacs, M. Suarez, J. Bacmeister, I.-S. Song, and A. Eichmann, 2012: The GEOS-5 atmospheric general circulation model: Mean climate and development from MERRA to Fortuna. NASA Tech. Memo. NASA/TM-2012-104606, Vol. 28, 117 pp. [Available online at https://gmao.gsfc.nasa.gov/pubs/docs/tm28.pdf.]

——, ——, ——, and ——, 2015: Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA-2. *Geosci. Model Dev.*, **8**, 1339–1356, doi:10.5194/gmd-8-1339-2015.

Moody, E. G., M. D. King, C. B. Schaaf, and S. Platnick, 2008: MODIS-derived spatially complete surface albedo products: Spatial and temporal pixel distribution and zonal averages. *J. Appl. Meteor. Climatol.*, **47**, 2879–2894, doi:10.1175/2008JAMC1795.1.

Mudryk, L. R., C. Derksen, P. J. Kushner, and R. Brown, 2015: Characterization of Northern Hemisphere snow water equivalent datasets, 1981–2010. *J. Climate*, **28**, 8037–8051, doi:10.1175/JCLI-D-15-0229.1.

Randles, C. A., and Coauthors, 2016: The MERRA-2 aerosol assimilation. NASA Tech. Memo. NASA/TM-2016-104606, Vol. 45, 132 pp. [Available online at https://gmao.gsfc.nasa.gov/pubs/docs/Randles887.pdf.]

Reichle, R. H., 2012: The MERRA-land data product. NASA GMAO Office Note 3 (version 1.2), 38 pp. [Available online at https://gmao.gsfc.nasa.gov/pubs/docs/Reichle541.pdf.]

——, R. D. Koster, G. J. M. De Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama, and A. Toure, 2011: Assessment and enhancement of MERRA land surface hydrology estimates. *J. Climate*, **24**, 6322–6338, doi:10.1175/JCLI-D-10-05033.1.

——, and Coauthors, 2016: Soil moisture active passive mission L4_SM data product assessment (version 2 validated release). NASA GMAO Office Note 12 (version 1.0), 55 pp.

——, Q. Liu, R. D. Koster, C. S. Draper, S. P. P. Mahanama, and G. S. Partyka, 2017: Land surface precipitation in MERRA-2. *J. Climate*, **30**, 1643–1664, doi:10.1175/JCLI-D-16-0570.1.

Reynolds, C. A., T. J. Jackson, and W. J. Rawls, 2000: Estimating soil water-holding capacities by linking the Food and Agriculture Organization Soil map of the world with global pedon databases and continuous pedotransfer functions. *Water Resour. Res.*, **36**, 3653–3662, doi:10.1029/2000WR900130.

Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, doi:10.1175/JCLI-D-11-00015.1.

Robertson, F. R., M. G. Bosilovich, J. B. Roberts, R. H. Reichle, R. Adler, L. Ricciardulli, W. Berg, and G. J. Huffman, 2014: Consistency of estimated global water cycle variations over the satellite era. *J. Climate*, **27**, 6135–6154, doi:10.1175/JCLI-D-13-00384.1.

Rodell, M., I. Velicogna, and J. Famiglietti, 2009: Satellite-based estimates of groundwater depletion in India. *Nature*, **460**, 999–1002, doi:10.1038/nature08238.

Rowlands, D. D., S. B. Luthcke, S. M. Klosko, F. G. R. Lemoine, D. S. Chinn, J. J. McCarthy, C. M. Cox, and O. B. Anderson, 2005: Resolving mass flux at high spatial and temporal resolution using GRACE intersatellite measurements. *Geophys. Res. Lett.*, **32**, L04310, doi:10.1029/2004GL021908.

Sakaguchi, K., and X. Zeng, 2009: Effects of soil wetness, plant litter, and under-canopy atmospheric stability on ground evaporation in the Community Land Model (CLM3.5). *J. Geophys. Res.*, **114**, D01107, doi:10.1029/2008JD010834.

Schaefer, G. L., M. H. Cosh, and T. J. Jackson, 2007: The USDA Natural Resources Conservation Service Soil Climate Analysis Network (SCAN). *J. Atmos. Oceanic Technol.*, **24**, 2073–2077, doi:10.1175/2007JTECHA930.1.

Smith, A., and Coauthors, 2012: The Murrumbidgee soil moisture monitoring network data set. *Water Resour. Res.*, **48**, W07701, doi:10.1029/2011WR010641.

Stieglitz, M., A. Ducharne, R. Koster, and M. Suarez, 2001: The impact of detailed snow physics on the simulation of snow cover and subsurface thermodynamics at continental scales. *J. Hydrometeor.*, **2**, 228–242, doi:10.1175/1525-7541(2001)002<0228:TIODSP>2.0.CO;2.

Swenson, S. C., 2012: GRACE monthly land water mass grids NETCDF RELEASE 5.0, version 5.0. Physical Oceanography Distributed Active Archive Center, accessed 31 May 2016, doi:10.5067/TELND-NC005.

——, and J. Wahr, 2006: Post-processing removal of correlated errors in GRACE data. *Geophys. Res. Lett.*, **33**, L08402, doi:10.1029/2005GL025285.

——, P. J.-F. Yeh, J. Wahr, and J. Famiglietti, 2006: A comparison of terrestrial water storage variations from GRACE with in situ measurements from Illinois. *Geophys. Res. Lett.*, **33**, L16401, doi:10.1029/2006GL026962.

Takacs, L. L., M. J. Suárez, and R. Todling, 2016: Maintaining atmospheric mass and water balance in reanalyses. *Quart. J. Roy. Meteor. Soc.*, **142**, 1565–1573, doi:10.1002/qj.2763.

USGS, 1996: Global 30 arc-second elevation (GTOPO30): Global topographic data. EROS Data Center Distributed Active Archive Center (EDC DAAC), accessed 1 July 2015. [Available online at https://lta.cr.usgs.gov/GTOPO30.]

——, 2000: Global land cover characteristics data base, version 2.0. Accessed 27 September 2016. [Available online at https://lta.cr.usgs.gov/glcc/globdoc2_0.]

Wahr, J., S. Swenson, and I. Velicogna, 2006: Accuracy of GRACE mass estimates. *Geophys. Res. Lett.*, **33**, L06401, doi:10.1029/2005GL025305.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press, 627 pp.

Wu, W.-S., R. J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905–2916, doi:10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2.

Xie, P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539–2558, doi:10.1175/1520-0477(1997)078<2539:GPAYMA>2.0.CO;2.

——, M. Chen, S. Yang, A. Yatagai, T. Hayasaka, Y. Fukushima, and C. Liu, 2007: A gauge-based analysis of daily precipitation over East Asia. *J. Hydrometeor.*, **8**, 607–626, doi:10.1175/JHM583.1.